**Proposal for Chinese Script Root Zone LGR**

# General Information

The purpose of this document is to give an overview of the proposed LGR in the XML format and the rationale behind the design.

It includes the discussions of script features together with the communities or language environments using it, the process and methodology generate the code point repertoire, code point variants and the WLE. It also indicates the issues need further coordination.

The first version was finished on June 13, 2016 and named for 20160613.

Integration Panel (IP) reviewed the first version and gave out two feedback documents as CGP-LGR20160613-Report-1-clean.docx and CGP-LGR20160613-Variant-Report\_0811-clean.docx.

CGP studied the feedback documents and generated the second version in September, 2016.

# Script and Languages Covered

## 1.1 Overview

Chinese characters are a sort of logogram used in the writing system of Chinese and some other Asian languages. They are called Hanzi in Chinese, Kanji in Japanese and Hanja in Korean.

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

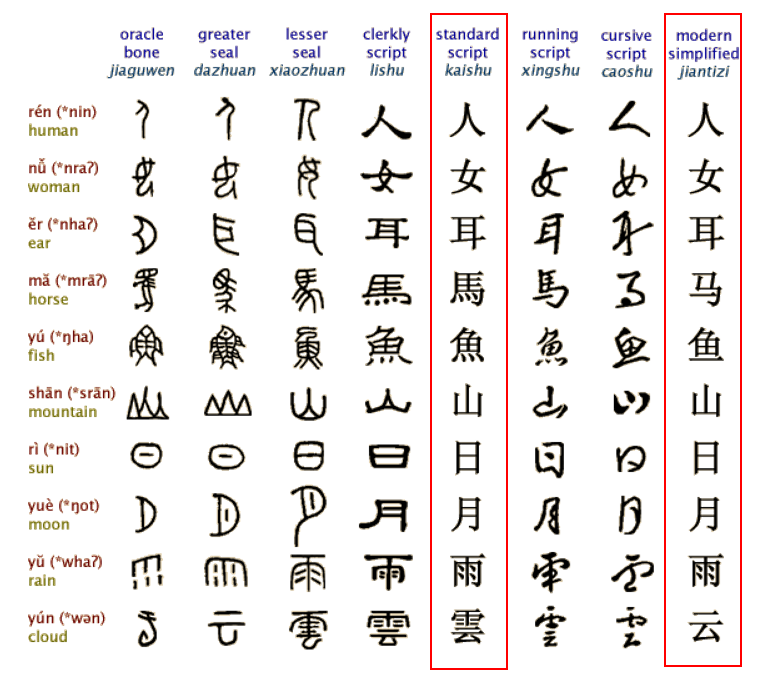


Figure 1: Evolution of Chinese Character

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

The Japanese kanji has been adopted for recording the Japanese language since the 5th century AD. Chinese words borrowed into Japanese could be written with Chinese characters, while Japanese words could be written using the character for a Chinese word of similar meaning. Finally, in Japanese, all three scripts (kanji, and the hiragana and katakana syllabaries) are used as main scripts.

The Chinese script was spreading to Korea together with Buddhism from the 2nd century BC to the 5th century AD. In times past, until the 15th century, in Korea, Literary Chinese was the dominant form of written communication, prior to the creation of Hangul, the Korean alphabet. In the modern Hangul-based Korean writing system, Chinese characters are no longer officially used to represent native morphemes, but still sometimes used in daily life.

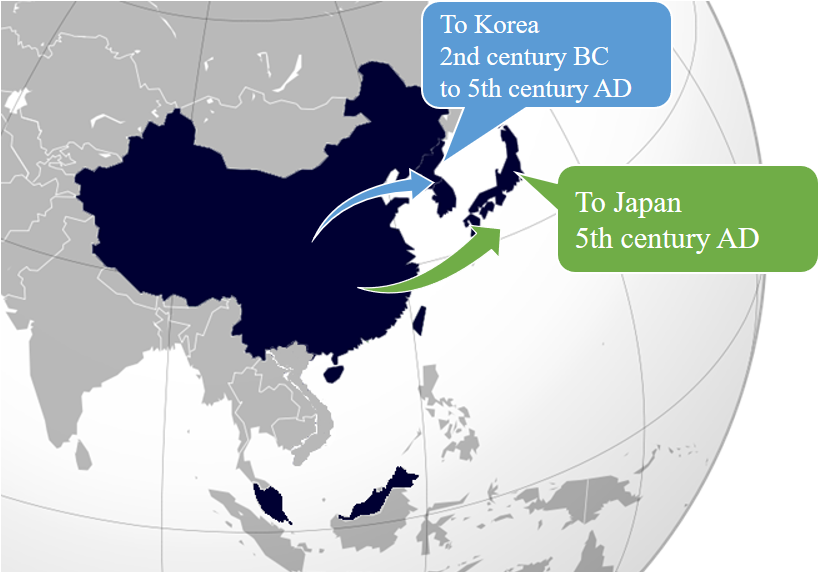
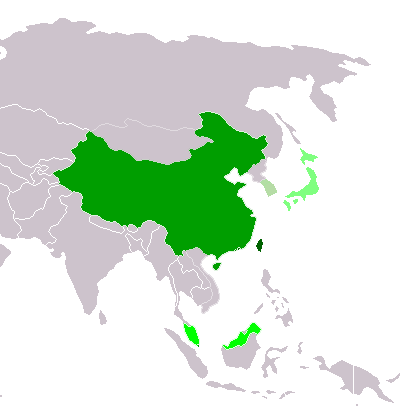


Figure 2: Chinese script spread to Japan and Korea

Historically, Chinese characters were also used in Mongolia and Vietnam, but not anymore. Accordingly, the Chinese Generation Panel does not take into account the usage of Chinese scripts in Mongolia and Vietnam.

## 1.2 Countries with Significant Usage for Chinese Script

Chinese script is used to write a diverse set of languages across East Asia and South East Asia. Countries and regions using Chinese script are depicted as follows:



|  |  |
| --- | --- |
|  | Traditional Chinese script used exclusively or almost exclusively  (Taiwan, Macau and Hong Kong) |
|  | Simplified Chinese script used exclusively or almost exclusively  (Mainland China and Singapore) |
|  | Simplified Chinese script used formally but Traditional script still used widely  (Malaysia) |
|  | Chinese script used with other systems of writing in the same language  Kanji (Japan) |
|  | Chinese script no longer officially used  Hanja (Republic of Korea) |

Figure 3: Countries using Chinese script

## 1.3 Target Script: Hani

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 ISO 15924, the script for Chinese Language is mainly defined in this specification:

ISO 15924 code: Hani

ISO 15924 no.: 500

English Name: Han (Hanzi, Kanji, Hanja)

Following ISO setting, CGP directly takes “Hani” as the Language Tag for Chinese.

## 1.4 Principal Languages using the Script

As shown in the following non-exhaustive table, Chinese, Japanese and Korean are three main languages using the Chinese script today but it does not imply that unlisted languages are less significant. For example, there are cases where a language may have a large population, but only a small part of it writes in Chinese script. Such languages are excluded from this list. For these language all ISO 639-3 available as “living” are included from <http://www-01.sil.org/ISO639-3/codes.asp>, which may refer to a macro or an individual language.

|  |  |  |  |
| --- | --- | --- | --- |
| **Language** | **ISO 15924 Code** | **Countries** | **Local Names of the Script** |
| Chinese | cdo, cjy, cmn, cpx, czh, czo, gan, hak, hsn, lzh, mnp, nan, wuu, yue, zho | China | 汉字 Hanzi |
| Japanese | jpn | Japan | 漢字 Kanji |
| Korean | kor | Korea | 한자 Hanja |

* Hanzi normally consists of two subsets, Simplified Chinese characters (Hans) and Traditional Chinese characters (Hant).
* Kanji is used in Japanese in addition to two other scripts (hiragana and katakana), together known as Jpan (ISO 15924 code).
* Hanja is used in Korean in addition to the Hangul script, together known as Kore (ISO 15924 code).

The relationship among Hanzi, Kanji and Hanja is as shown below, Hanzi (Hans & Hans), Kanji and Hanjia are both therefore covered by CGP.

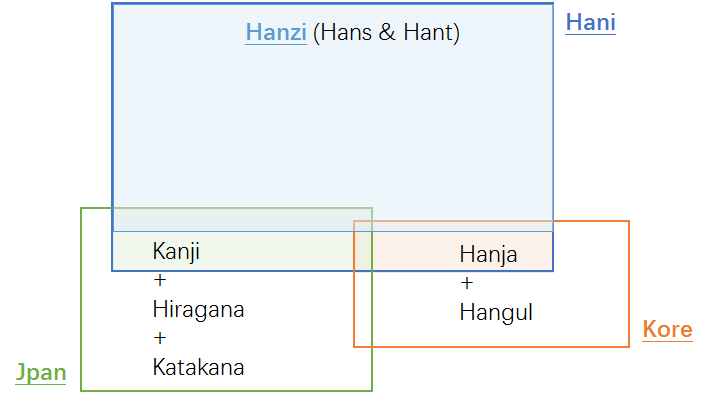


Figure 4: Hanzi, Kanji & Hanja

# Process for Developing the Proposal

## 2.1 Basis of the current work

In April 2004, the Joint Engineering Team (JET), a group composed of members of CNNIC, TWNIC, KRNIC, and JPNIC as well as other individual experts, produced RFC 3743, “Joint Engineering Team (JET) Guidelines for Internationalized Domain Names (IDN) Registration and Administration for Chinese, Japanese and Korean”, a guideline for zone administrators, including but not limited to registry operators and registrars and information for all domain names holders on the administration of domain names that contain characters drawn from the Chinese, Japanese, and Korean scripts. It includes concepts for variant handling, such as bundling, atomic IDL Packages, and reserved variants. It also defines a standard table as well as an algorithm to generate the preferred variant and reserved variants. The key mechanisms of this specification utilize a three-column table, called a Language Variant Table, for each language permitted to be registered in the zone.

Collectively, CDNC (Chinese Domain Name Consortium) has devised solutions to handle Chinese domain name variants, such as the bundling of Simplified Chinese (SC) and Traditional Chinese (TC) (“TC-SC Equivalence”) domain names — as defined by the JET in RFC 3743 (April 2004) and for the Chinese language as defined in RFC 4713 (October 2006) — and delegating the applied label, one preferred SC label and one preferred TC label to the same applicant. CDNC’s registration policy on handling TC-SC Equivalence is widely accepted. **CDNC IDN Table**, developed by many Chinese linguistic and domain name experts over the last 10 years is currently adopted by the Chinese, Taiwanese, Hong Kong, Macau and Singaporean governments, as well as by many new gTLD applicants. Over a decade of operating experience indicates CDNC’s TC-SC Equivalence solution is a market-proven successful practice for handling Chinese variants in domain names.

Meanwhile, the registry of .asia as well as the member of CDNC, dotAsia, extended CDNC IDN Table by importing characters from HKSCS (Hong Kong Supplementary Character Set) and Singapore set, developed its own IDN table under the framework of CDNC rules, to cover the needs of Hong Kong and Singapore local community.

There has already been a detailed analysis of Chinese script done by the community in an earlier phase of the LGR program, which resulted in a **Chinese Case Study Team Report** (<https://archive.icann.org/en/topics/new-gtlds/chinese-vip-issues-report-03oct11-en.pdf>).

All these above previous efforts made by the Chinese script community have been used as a basis for the current work, especially the Chinese Study Report and RFC 4713, in addition to other literature and the expertise available in the current task force.

## 2.2 Team Diversity

The current work is undertaken by experts from CDNC, who represent the Chinese language ccTLDs to a large extent, as well as experts from a variety of backgrounds.

Geographically, the CGP has members from Chinese language regions across east Asia, including China mainland, Taiwan, Hong Kong, Macau, Singapore, Malaysia, as well as members from Europe and North America, totally 23 members belonging to 10 countries/regions listed in Appendix A.

CGP consists of members with a diverse set of disciplines and very different perspectives. The members represent national and regional policy makers, technical community directly working with the DNS, security and law enforcement community, academia (technical and linguistic), and the members with experience with local language computing using Unicode and specifically IDNs.

Besides, CGP is pleased to have **Edmon CHUNG**, CEO of dotAsia and Co-Chair of Universal Acceptance Steering Group, as the IDN advisor.

## 2.3 Work Process

The work has been carried out since September 2014, when the group formed to put forward a “proposal for generation panel for Chinese script label generation ruleset for the root zone”. Since then, the group has had fortnightly conference calls, as well as two face-to-face meetings along with the CDNC annual meeting, in July 2015 and March 2016. In addition, the group has been actively engaged over email, through the public mailing list of the task force.

The group also maintains frequent communication with JGP and KGP, to coordinate the Chinese code points and variant characters among three parties. Three Parties held jointly two face-to-face meetings, in May 2015 and March 2016, and hold successive CJK joint session in ICANN meeting since ICANN 51 Los Angeles.

The work has been structured to make the following steps:

* Define and finalize the code point repertoire

In the range of MSR, CDNC and most CGP members urged to add CDNC characters into CGP repertoire as much as possible, to reach consistency between the CDNC SLD operation and future TLD operation.

In addition to CDNC IDN Table, there are some other character sets taken into account, including dotAsia IDN Table, Normalized Hanzi List for Common Use (NHCU) published by China State Council, and IICORE (International Ideographs Core).

JGP repertoire and KGP repertoire are also referred to.

* Define and finalize the code point variant sets

CDNC has given an industry-proven variant set in CDNC IDN table. Following CDNC rules, dotAsia extended CDNC repertoire and variant set to meet the requirement from Hong Kong and Singapore local community. Basically, CGP would adopt CDNC variant rule directly before the coordination with J, K and IP.

CGP recognizes that different panels (C, J and K) have different views on the variants corresponding to the same Chinese character, some CGP variant mappings conflict with KGP and JGP’s perception and practice. CGP was working closely with JGP & KGP to make necessary compromises to reach a consensus for all three parties and meet the IP’s requirement that “The variant mappings must agree for the same code point for all LGRs”.

* Define and finalize whole label evaluation ruleset

The CGP WLE follows the spirit of CDNC ruleset, “TC-SC equivalence”, which assigns all variant labels to the same applicant, while allocating the original applied label as well as only preferred SC label(s) and preferred TC label(s), generally no more than three labels, and blocks all other labels.

CGP also acknowledges that some multiple preferred variant mappings work for SLD but may overproduce allocatable labels in root zone. CGP will work together with J, K and IP to devise an ideal solution to set applicant preferred labels allocatable as well as to limit the allocatable labels amount within a reasonable number (like three).

* Create XML LGR for Chinese script LGR proposal

Considering the fact that the coordination on repertoire, variant mappings and WLE with J, K and IP are still in the progress, the CGP work will be carried out in a fast iteration model as indicated in the following figure:

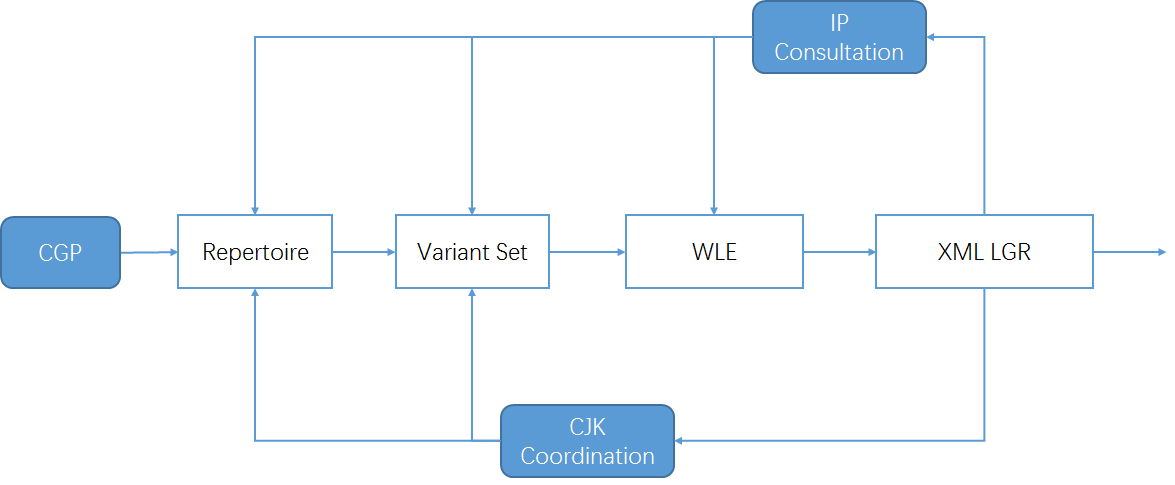


Figure 5: Iteration model of CGP work process

# Code point repertoire

## 3.1 Basic character set

In 2004, according to RFC 3743 and RFC 4713, the Chinese Domain Name Consortium (CDNC) submitted to IANA a unified Chinese Character Set for domain name registration, building up mapping relationships between any given simplified character, its traditional character(s) and its variant(s). The CDNC Character Set has been used for second level domain (SLD) name registration under .CN, .TW, .HK and etc. since 2002.

In the early work, CGP tried to reduce the character number of CDNC IDN table, from 19000+ to 12000+, into a smaller subset, hoping the reduction will help decrease the [computational](http://cn.bing.com/dict/search?q=computational&FORM=BDVSP6&mkt=zh-cn) [complexity](http://cn.bing.com/dict/search?q=complexity&FORM=BDVSP6&mkt=zh-cn) and speed up the coordination work with J & K. But soon after that, CGP realized it is the variant mapping rules, not the repertoire size that directly affect the computational complexity. The number of disputed variant mapping characters among CJK is only 258, which means the issue could be addressed case by case based on finite statistic work. Moreover, considering the TLD market acceptance of the existing CDNC IDN Table and the continuity of registrars and applicants experience, urged by CDNC and most CGP members, CGP finally decided to accept all CDNC Characters as the basic character set of CGP repertoire, to reach consistency between the CDNC SLD operation and future TLD operation.

## Repertoire formation process

* In March 2012, CDNC reviewed and published its IDN Table for Chinese domain name registration as <http://www.cdnc.org/gb/research/file/CDNC_unicode.txt>, including 19557 code points (37 ASCII code point and 19520 Chinese characters).   
  19520 Hanzi ideographs from two Unicode blocks: CJK UNIFIED IDEOGRAPHS and CJK UNIFIED IDEOGRAPHS EXTENSION A. All these **19520** characters are included in MSR.
* In October 2015, CDNC published the latest version of IDN Table, 41 new Chinese characters were added into the character set as requested by HKIRC on behalf of Hong Kong local community in 2013 and 2015, increasing the number of Chinese characters to 19561, but two of them (3A5C㩜 and 58B5墵) are out of scope of MSR2, which means only 19559 are included in MSR2.

All **19561** code points form the basic set of CGP repertoire. (CGP R0=CDNC IDN Table)

* dotAsia extended CDNC IDN Table, adding 163 code points, of which 156 are part of HKSCS included in the IICORE collection, 4 are GS (Singapore Characters), and the remaining 3 are part of various other Chinese sources that are necessary to insure full transitivity in variant processing, made up the .asia repertoire of 19683 code points. (<https://www.iana.org/domains/idn-tables/tables/asia_zh_1.1.txt>)

39 of 163 are already included in CDNC IDN Table 2015, the rest 124 ones extended CGP repertoire up to **19685** code points (CGP R1, covering whole dotAsia IDN Table).

CGP R1 covers 7 Hanzi characters not included in CGP LGR 20160613.

|  |  |
| --- | --- |
| 39DB | 㧛 |
| 3BA3 | 㮣 |
| 43D3 | 䏓 |
| 4443 | 䑃 |
| 4882 | 䢂 |
| 4C9D | 䲝 |
| 4C9E | 䲞 |

CGP has to point out that, unlike all the other Hanzi code points (U+3000~U+9FFF) in Basic Multilingual Plane, 62 code points in dotAsia IDN　table from Supplementary Ideographic Plane are hard to be displayed in many operation and application systems.

|  |  |  |  |
| --- | --- | --- | --- |
| 2070E |  | 210C9 |  |
| 20731 |  | 211D9 |  |
| 20779 |  | 220C7 |  |
| 20C53 |  | 227B5 |  |
| 20C78 |  | 22AD5 |  |
| 20C96 |  | 22B43 |  |
| 20CCF |  | 22BCA |  |
| 20CD5 |  | 22C51 |  |
| 20D15 |  | 22C55 |  |
| 20D7C |  | 22CC2 |  |
| 20D7F |  | 22D08 |  |
| 20E0E |  | 22D4C |  |
| 20E0F |  | 22D67 |  |
| 20E77 |  | 22EB3 |  |
| 20E9D |  | 23CB7 |  |
| 20EA2 |  | 244D3 |  |
| 20ED7 |  | 24DB8 |  |
| 20EF9 |  | 24DEA |  |
| 20EFA |  | 2512B |  |
| 20F2D |  | 26258 |  |
| 20F2E |  | 267CC |  |
| 20F4C |  | 269F2 |  |
| 20FB4 |  | 269FA |  |
| 20FBC |  | 27A3E |  |
| 20FEA |  | 2815D |  |
| 2105C |  | 28207 |  |
| 2106F |  | 282E2 |  |
| 21075 |  | 28CCA |  |
| 21076 |  | 28CCD |  |
| 2107B |  | 28CD2 |  |
| 210C1 |  | 29D98 |  |

* In 2013, the China's State Council published Normalized Hanzi List for Common Use (NHCU), CGP studied the NHCU, found 18 characters out of repertoire and fall in the range of MSR as well. CGP imported these 18 characters into the repertoire, making the total number **19703**. (CGP R2)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | NHCU | IICORE | | | | | | | |
| G | T | J | H | K | M | KP | S |
| 48BC | 䢼 | N (normalized) | G3D |  |  |  |  |  |  | C |
| 732F | 猯 | N |  |  |  |  |  |  |  |  |
| 9EB9 | 麹 | N |  |  | J1A |  |  |  |  | A |
| 5227 | 刧 | V (variant to N) |  |  |  |  |  |  |  |  |
| 524F | 剏 | V |  |  |  |  |  |  |  |  |
| 6060 | 恠 | V |  |  |  |  |  |  |  |  |
| 74A2 | 璢 | V |  |  |  |  |  |  |  |  |
| 750E | 甎 | V |  |  |  |  |  |  |  |  |
| 754A | 畊 | V |  |  |  |  |  |  |  |  |
| 7ADA | 竚 | V |  |  |  |  |  |  |  |  |
| 8262 | 艢 | V |  |  |  |  |  |  |  |  |
| 88B5 | 袵 | V |  |  |  |  |  |  |  |  |
| 894D | 襍 | V |  |  |  |  |  |  |  |  |
| 8B0C | 謌 | V |  |  |  |  |  |  |  |  |
| 8F19 | 輙 | V |  |  |  |  |  |  |  |  |
| 945A | 鑚 | V |  |  | J1A |  |  |  |  | C |
| 984B | 顋 | V |  |  |  |  |  |  |  |  |
| 9DC0 | 鷀 | V |  |  |  |  |  |  |  |  |

* Furthermore, IICORE, JGP repertoire and KGP repertoire were studied.  
  In early 2000s, CDNC experts focused on modern frequently used characters, excluded some IICORE characters from CDNC IDN Table (CGP R0). Those characters might be included in JGP repertoire or KGP repertoire, and could be variants of current CGP characters.  
  To ensure that CGP repertoire will not bring any confusion or conflict to global Chinese character users and applicants at root level, CGP reviewed 99 IICORE characters which are overlapped by MSR but not covered by CGP R2, found 43 characters included in JGP repertoire (version 20160316, Appendix C) and KGP repertoire (version 20160820, Appendix D) have variant mapping relationship with CGP R2. [see section 4.3]

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | JGP | KGP | IICORE | | | | | | | |
| G | T | J | H | K | M | KP | S |
| 3960 | 㥠 |  | K |  | T3D | J1A |  |  |  |  | A |
| 4FAD | 侭 | J |  |  |  | J1A |  |  |  |  | C |
| 51E6 | 処 | J |  |  |  | J1A |  |  |  |  | A |
| 56A2 | 嚢 | J |  |  |  |  |  |  |  |  | A |
| 61F4 | 懴 | J |  |  |  | J1A |  |  |  |  | A |
| 6442 | 摂 | J |  |  |  | J1A |  |  |  |  | A |
| 663B | 昻 |  | K |  |  | J1A |  |  |  |  | A |
| 685C | 桜 | J |  |  |  | J1A |  |  |  |  | A |
| 685F | 桟 | J |  |  |  | J1A |  |  |  |  | A |
| 6D9C | 涜 | J |  |  |  | J1A |  |  |  |  | C |
| 6E8C | 溌 | J |  |  |  | J1A |  |  |  |  | A |
| 731F | 猟 | J |  |  |  | J1A |  |  |  |  | A |
| 784F | 硏 |  | K |  |  |  |  | K3D |  |  | C |
| 7A36 | 稶 |  | K |  |  | J1A |  |  |  |  | A |
| 7B86 | 箆 | J |  |  |  | J1A |  |  |  |  | C |
| 7C14 | 簔 | J |  |  |  | J1A |  |  |  |  | A |
| 7D9A | 続 | J |  |  |  |  |  |  |  |  | A |
| 7E4A | 繊 | J |  |  |  | J1A |  |  |  |  | A |
| 7E4B | 繋 | J |  |  |  | J1A |  |  |  |  | A |
| 8133 | 脳 | J |  |  |  |  |  | K0A |  | P0A | A |
| 81D3 | 臓 | J |  |  |  | J1A |  |  |  |  | C |
| 8217 | 舗 | J |  |  |  | J1A |  |  |  |  | A |
| 839F | 莟 | J |  |  |  | J1A |  |  |  |  | A |
| 83B5 | 莵 | J |  |  |  | J1A |  |  |  |  | A |
| 86CD | 蛍 | J |  |  |  | J1A |  |  |  |  | C |
| 8E99 | 躙 | J |  |  |  | J1A |  |  |  |  | A |
| 9039 | 逹 | J |  |  |  | J1A |  |  |  |  | A |
| 91A4 | 醤 | J |  |  |  | J1A |  |  |  |  | C |
| 91C8 | 釈 | J |  |  |  | J1A |  |  |  |  | A |
| 9271 | 鉱 | J |  |  |  |  |  | K0A |  | P0A | A |
| 9421 | 鐡 | J |  |  |  | J1A |  |  |  |  | A |
| 967A | 険 | J |  |  |  | J1A |  |  |  |  | A |
| 96B2 | 隲 | J |  |  |  | J1A |  |  |  |  | A |
| 982C | 頬 | J |  |  |  | J1A |  |  |  |  | C |
| 98EE | 飮 | J | K |  |  | J1A |  |  |  |  | A |
| 9A12 | 騒 | J |  |  |  | J1A |  |  |  |  | A |
| 9A13 | 験 | J |  |  |  | J1A |  |  |  |  | A |
| 9A28 | 騨 | J |  |  |  |  |  | K0A |  | P0A | A |
| 9C2E | 鰮 | J |  |  |  |  |  | K0A |  | P0A | A |
| 9D0E | 鴎 | J |  |  |  | J1A |  |  |  |  | A |
| 9D2C | 鴬 | J |  |  |  | J1A |  |  |  |  | C |
| 9D8F | 鶏 | J |  |  |  | J1A |  |  |  |  | C |
| 9EBA | 麺 | J |  |  |  | J1A |  |  |  |  | A |

CGP then imported these 43 characters into repertoire to make sure that “all variant mappings must agree for all three panels”, raising the total number to **19746.** (CGP R3)

The intersections between CGP repertoire and other Chinese character sets are illustrated as the following figure:

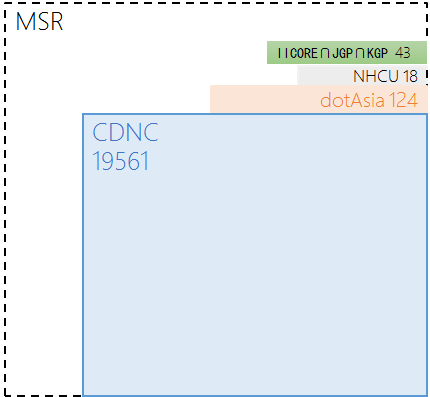


Figure 6: CGP R3 repertoire components

## 3.2 Repertoire coordination with JGP and KGP

* For 6358 Hani/Kanji characters in JGP repertoire (version 20160316, Appendix C), there are 6267 overlapped characters in CGP R3 repertoire. Beyond that, CGP will treat the others as Japanese UNIQUE Kanji characters and will not seek to borrow any other characters from JGP.
* For 4819 Hani/Hanja characters in KGP repertoire (version 20160820, Appendix D), there are 4809 characters are included in CGP R3 repertoire. Beyond that, CGP will treat the others as Korean UNIQUE Hanja characters and will not seek to borrow any other characters from KGP.

The relationship among CGP repertoire, JGP repertoire and KGP repertoire is illustrated as the following figure:

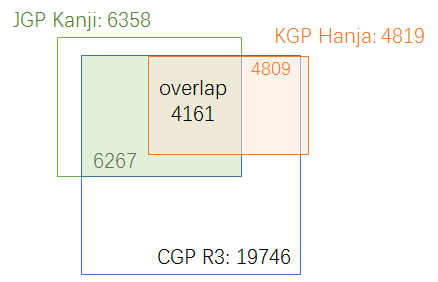


Figure 7：CGP Hanzi set, JGP Kanji set and KGP Hanja set

Given that JGP and KGP might disagree with some variant character groups and mappings, in order to reach a consensus for all the three parties, one of the compromise tools is to remove the controversial characters out of repertoire. So before the final whole label ruleset is made, CGP will keep coordinating with JGP and KGP, modifying CGP repertoire where necessary.

# Code point variants

## 4.1 Variant definition in CGP

In Chinese language, there are two types of variants:

The first type is created by regional variations in the standard writing system. There are now two common writing systems: Simplified Chinese and Traditional Chinese. Both writing systems use different subsets of the same Unicode Han script, but they are not mutually exclusive to each other.

The second type is the generic variant. Several Chinese characters are visually different in forms, but treated equally with universal interchangeability. This relationship of interchangeability is much stronger than the relationship between the Traditional and Simplified forms.

In Chinese Case Study Team Report mentioned in 2.1, 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.”**

This understanding and variants mapping rule has been reflected in the CDNC IDN Table, and followed by current CGP LGR document.

In alignment with RFC 4713 and CDNC practice, generally, every code point in CGP repertoire has its preferred simplified variant(s), preferred traditional variant(s), and reserved variant(s). In some case, a code point has reflexive preferred variant. In some other, a code point has no reserved variant.

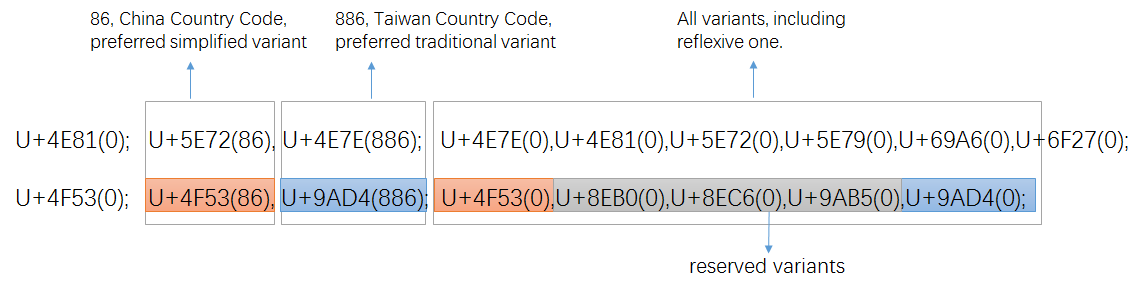


Figure 8: variant setting in CDNC IDN Table

Once transformed into XML-format (draft-davies-idntables-10, Representing Label Gneration Rulesets using XML, <https://datatracker.ietf.org/doc/draft-davies-idntables/>), all preferred variant char(s) are “allocatable", all reserved variant char(s) are “blocked”, with sub-types as:

|  |  |  |
| --- | --- | --- |
| Sub-Type | Type | Comment |
| “simp” | Allocatable | preferred simplified variant char; |
| “r-simp” | Allocatable | reflexive preferred simplified variant char; |
| “trad” | Allocatable | preferred traditional variant char |
| “r-trad” | Allocatable | reflexive preferred traditional variant char |
| “both” | Allocatable | preferred simplified and traditional variant chars are the same |
| “r-both” | Allocatable | reflexive preferred simp and trad variant chars are the same |
| “r-neither” | Blocked | Non-allocatable reflexive/original char |
| “blocked” | Blocked | Non-allocatable variant char |

In alignment of XML rules, the two variant mappings in Figure8 will be transformed into the following text:

<char cp="4F53" tag="sc:Hani" >

<var cp="4F53" type="r-simp" comment="identity" />

<var cp="8EB0" type="blocked" />

<var cp="8EC6" type="blocked" />

<var cp="9AB5" type="blocked" />

<var cp="9AD4" type="traded" />

</char>

<char cp="4E81" tag="sc:Hani" >

<var cp="4E7E" type="trad" />

<var cp="4E81" type="r-neither" comment="identity" />

<var cp="5E72" type="simp" />

<var cp="5E79" type="blocked" />

<var cp="69A6" type="blocked" />

<var cp="6F27" type="blocked" />

</char>

Note: To reduce the number of allocatable labels in WLE, besides the above sub-types, more new sub-types were created and introduced in Section 4.6, to eliminate multiple variant mappings.

## 4.2 CJK Coordination framework

Some Kanji characters are in a simplified form (called the “new character form”), derived from the traditional imported form (called the “old character form”). In Japanese language environment and writing system, it is appropriate to distinguish NEW and OLD forms as different and independent characters instead of pure 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 as in a bilateral discussion in IETF Dallas meeting, March 2015,

Hanja characters are no longer used in official documents (A law 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), but still sometimes are used by a few Korean people in daily life. However, in August 2016, KGP generated its first version of LGR, including 4819 Hanja characters with 95 variant characters and 47 variant groups (Appendix E). Moreover, KGP raised 258 Hanja/Hanzi characters whose variant mappings in CGP LGR are NOT acceptable for KGP (Appendix F).

A coordination scheme among different panels is needed in that we expect unified Chinese script generation rules in the DNS root zone. During the CDNC meeting in Shanghai (May, 2014), IP proposed the basic principles of the coordination scheme:

* Each CJK panel creates an LGR and each LGR includes a repertoire and variants.
* If an LGR includes Han characters, the variant mappings must agree for all three panels.
* The variant types may be different (blocked or allocatable), the variant types do not have to be agreed on across LGRs.

Based on the principles above, CGP, JGP and KGP started coordination work since IETF Dallas meeting 2015, trying to define a unified variant mapping table for Chinese scripts, then define each party’s variant types/sub-types (e.g., allocatable or blocked) for characters contained in this table. According to the consensus in IETF Dallas meeting, JGP initiated a work called “CJK Integration Procedure” as follows:

Step 1: Each CJK GP generates its own LGR (hereinafter, LGR-alpha)

Step 2: CJK GPs collectively generate a merged table of each LGR-alpha (hereinafter, LGR-M)

Step3: Each CJK GP extracts its original repertoire with integrated variants from LGR-M.

Step 4: Each CJK GP adds “Out of Repertoire” code points for symmetry.

Step 5: Each CJK GP merge WLE in LGR-alpha into one.

Step 6: Each CJK GP generates integrated LGR (hereinafter, LGR-beta).

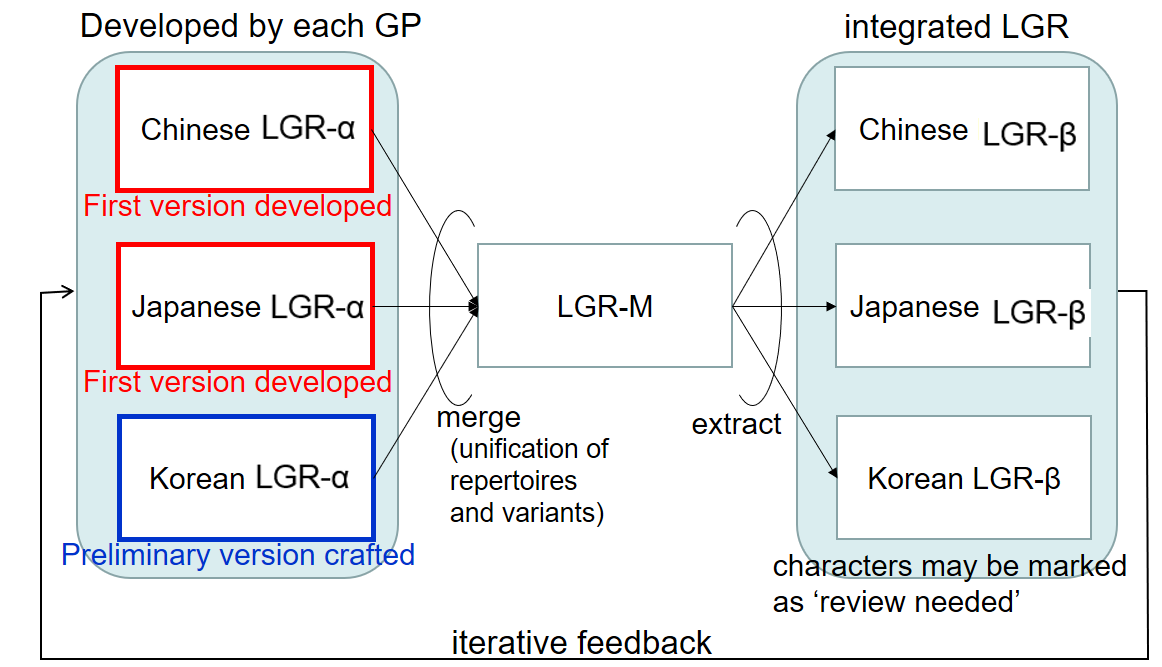


Figure9: Framework of CJK LGR integration for Han characters, by JGP

## 4.3 Variants and variant mappings review

Under the above coordination framework, CDNC started character review work in August and September 2015. In early 2000s, when drafting the IDN table, CDNC experts focused on modern frequently used characters and excluded some IICORE characters from CDNC IDN Table (CGP R0). Those missing characters might be included in JGP repertoire or KGP repertoire, and could be variants of previous CGP characters. To ensure that CGP repertoire will not bring any confusion or conflict to global Chinese character users and applicants at root level, CGP & CDNC held joint meetings and invited linguistic experts from China mainland, Taiwan and Hong Kong to review 172 Hanzi characters not included in CGP R1.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **Hong Kong** | **NHCU** | **IICORE** | | | | | | | | **JGP** | **KGP** | **.asia** |
| **G** | **T** | **J** | **H** | **K** | **M** | **KP** | **S** |
| 34E4 | 㓤 |  |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 3577 | 㕷 |  |  |  | T3B |  |  |  |  |  | C |  |  | A |
| 35A1 | 㖡 |  |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 35AD | 㖭 |  |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 35BF | 㖿 |  |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 35CE | 㗎 |  |  |  |  |  | H1F |  | M1F |  | B |  |  | A |
| 35F3 | 㗳 |  |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 35FE | 㗾 |  |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 3960 | 㥠 |  |  |  |  |  |  | K3D |  |  | C |  | K |  |
| 39F8 | 㧸 |  |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 39FE | 㧾 |  |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 3A18 | 㨘 |  |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 3A52 | 㩒 |  |  |  |  |  | H1F |  | M1F |  | B |  |  | A |
| 3A67 | 㩧 |  |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 3B39 | 㬹 |  |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 3DE7 | 㷧 |  |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 3DEB | 㷫 |  |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 3E74 | 㹴 |  |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 3ED0 | 㻐 |  |  |  |  |  |  |  |  | P0A | C |  | K | A |
| 4065 | 䁥 |  |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 406A | 䁪 |  |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 40BB | 䂻 |  |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 40DF | 䃟 |  |  |  |  |  | H1E |  |  |  | C |  |  | A |
| 4137 | 䄷 |  |  |  |  |  |  | K3D |  |  | C |  | K |  |
| 44EA | 䓪 |  |  |  |  |  | H1D |  |  |  | C |  |  | A |
| 4606 | 䘆 |  |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 47F4 | 䟴 |  |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 48B5 | 䢵 |  |  | G5D |  |  |  |  |  |  | C |  |  |  |
| 48BC | 䢼 |  | N | G3D |  |  |  |  |  |  | C |  |  |  |
| 48C5 | 䣅 |  |  | G3D |  |  |  |  |  |  | C |  |  |  |
| 48D3 | 䣓 |  |  | G3D |  |  |  |  |  |  | C |  |  |  |
| 49D1 | 䧑 |  |  | G9D |  |  |  |  |  |  | C |  |  |  |
| 4A12 | 䨒 |  |  |  |  |  |  | K3D |  |  | C |  | K |  |
| 4AB8 | 䪸 |  |  |  |  |  |  | K3D |  |  | C |  | K | A |
| 4C7D | 䱽 |  |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 4C81 | 䲁 |  | V |  | T4B |  |  |  |  |  | C |  |  | A |
| 4C85 | 䲅 |  |  |  | T4B |  |  |  |  |  | C |  |  | A |
| 4CB3 | 䲳 |  |  |  | T3B |  |  |  |  |  | C |  |  |  |
| 4D08 | 䴈 |  |  |  | T4B |  |  |  |  |  | C |  |  |  |
| 4E55 | 乕 |  |  |  |  | J1A |  |  |  |  | C | J |  |  |
| 4EEE | 仮 | HK2015 |  |  |  | J1A |  |  |  |  | A | J |  | A |
| 4FAD | 侭 |  |  |  |  |  |  |  |  |  | A | J |  |  |
| 51B4 | 冴 | HK2015 |  |  |  | J1A |  |  |  |  | A | J |  | A |
| 51E6 | 処 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 5227 | 刧 |  | V |  |  |  |  |  |  |  |  | J |  |  |
| 524F | 剏 |  | V |  |  |  |  |  |  |  |  | J |  |  |
| 5271 | 剱 |  |  |  |  | J1A |  |  |  |  | C | J |  |  |
| 5368 | 卨 |  |  |  |  |  |  | K0A |  | P0A | A |  | K |  |
| 5605 | 嘅 | HK2015 | V |  |  |  | H1F |  | M1F |  | B |  |  | A |
| 5689 | 嚉 | HK2015 |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 56A2 | 嚢 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 576E | 坮 |  |  |  |  |  |  | K0A |  | P0A | A |  | K |  |
| 57DE | 埞 | HK2015 |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 5817 | 堗 |  |  |  |  |  |  |  |  | P0A | C |  | K |  |
| 5841 | 塁 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 58CC | 壌 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 5BFE | 対 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 5C02 | 専 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 5CC0 | 峀 |  |  |  |  |  |  | K0A |  | P0A | A |  | K |  |
| 5D5C | 嵜 |  |  |  |  | J1A |  |  |  |  | C | J |  |  |
| 5E3F | 帿 |  |  |  |  |  |  | K0A |  | P0A | A |  | K |  |
| 5F10 | 弐 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 6060 | 恠 |  | V |  |  |  |  |  |  |  |  | J |  |  |
| 60A9 | 悩 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 60E3 | 惣 | HK2013 |  |  |  | J1A |  |  |  |  | A | J |  | A |
| 61F4 | 懴 |  |  |  |  | J1A |  |  |  |  | C | J |  |  |
| 62A6 | 抦 | HK2015 |  |  |  |  | H1F |  |  |  | C |  |  | A |
| 6335 | 挵 | HK2013 | V |  | T3B |  |  |  |  |  | C |  |  | A |
| 633F | 挿 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 637F | 捿 | HK2015 |  |  |  |  |  | K0A |  | P0A | A |  | K | A |
| 63BB | 掻 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 6442 | 摂 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 656D | 敭 | HK2013 | V |  |  |  |  | K0A |  | P0A | A |  | K | A |
| 65FF | 旿 | HK2015 | N |  |  |  |  | K0A |  | P0A | A |  | K | A |
| 663B | 昻 |  |  |  |  |  |  | K0A |  | P0A | A |  | K |  |
| 663F | 昿 |  |  |  |  | J1A |  |  |  |  | C | J |  |  |
| 6667 | 晧 | HK2013 |  |  |  | J1A |  | K0A |  | P0A | A | J | K | A |
| 66FD | 曽 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 681E | 栞 | HK2013 | V |  |  | J1A |  | K1C |  |  | B | J | K | A |
| 685C | 桜 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 685F | 桟 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 6AAA | 檪 |  |  |  |  | J1A |  |  |  |  | C | J |  |  |
| 6C17 | 気 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 6D9C | 涜 |  |  |  |  |  |  |  |  |  | A | J |  |  |
| 6E13 | 渓 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 6E7A | 湺 |  |  |  |  |  |  | K0A |  | P0A | A |  | K |  |
| 6E8C | 溌 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 701E | 瀞 | HK2015 |  |  | T3G | J1A |  | K0A |  | P0A | A | J | K | A |
| 7114 | 焔 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 713C | 焼 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 7155 | 煕 |  |  |  |  | J1A |  |  |  |  | C | J |  |  |
| 72A0 | 犠 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 731F | 猟 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 732F | 猯 |  | N |  |  |  |  |  |  |  |  | J |  |  |
| 7363 | 獣 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 7460 | 瑠 | HK2015 | V |  | T3D | J1A |  | K0A |  | P0A | A | J | K | A |
| 74A2 | 璢 |  | V |  |  |  |  |  |  |  |  | J |  |  |
| 74C8 | 瓈 | HK2015 | V |  | T3G |  |  |  |  |  | C |  |  | A |
| 750E | 甎 |  | V |  |  |  |  |  |  |  |  | J |  |  |
| 7534 | 甴 | HK2015 |  |  |  |  | H1F |  | M1C |  | B |  |  | A |
| 754A | 畊 |  | V |  |  |  |  |  |  |  |  | J |  |  |
| 7573 | 畳 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 757A | 畺 | HK2015 |  |  |  |  |  | K0A |  | P0A | A |  | K | A |
| 75E9 | 痩 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 7807 | 砇 |  |  |  |  |  |  | K3D |  |  | C |  | K |  |
| 783F | 砿 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 784F | 硏 |  |  |  |  |  |  | K0A |  | P0A | A |  | K |  |
| 7A36 | 稶 |  |  |  |  |  |  | K0A |  | P0A | A |  | K |  |
| 7A4F | 穏 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 7A63 | 穣 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 7AC3 | 竃 | HK2015 |  |  |  | J1A |  |  |  |  | A | J |  | A |
| 7AD7 | 竗 |  |  |  |  |  |  | K0A |  | P0A | A |  | K |  |
| 7ADA | 竚 |  | V |  |  |  |  |  |  |  |  | J |  |  |
| 7B86 | 箆 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 7C14 | 簔 |  |  |  |  | J1A |  |  |  |  | C | J |  |  |
| 7C4F | 籏 |  |  |  |  | J1A |  |  |  |  | C | J |  |  |
| 7D9A | 続 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 7E4A | 繊 |  |  |  | T3D | J1A |  |  |  |  | A | J |  |  |
| 7E4B | 繋 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 8133 | 脳 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 81D3 | 臓 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 8217 | 舗 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 8262 | 艢 |  | V |  |  |  |  |  |  |  |  | J |  |  |
| 839F | 莟 |  |  |  |  | J1A |  |  |  |  | C | J |  |  |
| 83B5 | 莵 |  |  |  |  | J1A |  |  |  |  | C | J |  |  |
| 8420 | 萠 | HK2015 |  |  |  | J1A |  |  |  |  | C | J |  | A |
| 86CD | 蛍 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 874B | 蝋 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 877F | 蝿 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 88B5 | 袵 |  | V |  |  |  |  |  |  |  |  | J |  |  |
| 894D | 襍 |  | V |  |  |  |  |  |  |  |  | J |  |  |
| 8A33 | 訳 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 8AAD | 読 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 8B0C | 謌 |  | V |  |  |  |  |  |  |  |  | J |  |  |
| 8B72 | 譲 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 8E99 | 躙 |  |  |  |  | J1A |  |  |  |  | C | J |  |  |
| 8F0C | 輌 |  |  |  |  | J1A |  |  |  |  | C | J |  |  |
| 8F19 | 輙 |  | V |  |  |  |  |  |  |  |  | J |  |  |
| 9039 | 逹 |  |  |  |  | J1A |  |  |  |  | C | J |  |  |
| 9197 | 醗 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 91A4 | 醤 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 91B8 | 醸 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 91C8 | 釈 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 9244 | 鉄 | HK2015 |  |  |  | J1A |  |  |  |  | A | J |  | A |
| 9271 | 鉱 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 932C | 錬 | HK2015 |  |  |  | J1A |  |  |  |  | A | J |  | A |
| 9421 | 鐡 |  |  |  |  | J1A |  |  |  |  | C | J |  |  |
| 945A | 鑚 |  | V |  |  | J1A |  |  |  |  | C | J |  |  |
| 9665 | 陥 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 967A | 険 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 96B2 | 隲 |  |  |  |  | J1A |  |  |  |  | C | J |  |  |
| 974D | 靍 |  |  |  |  | J1A |  |  |  |  | C |  |  |  |
| 9771 | 靱 | HK2015 | V |  |  | J1A |  |  |  |  | C | J |  | A |
| 982C | 頬 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 984B | 顋 |  | V |  |  |  |  |  |  |  |  | J |  |  |
| 98C7 | 飇 | HK2015 |  |  |  |  |  | K0A |  | P0A | A |  | K | A |
| 98E1 | 飡 | HK2015 |  |  |  |  |  | K0A |  | P0A | A |  | K | A |
| 98EE | 飮 |  |  |  |  |  |  | K0A |  | P0A | A | J | K |  |
| 99C5 | 駅 | HK2013 |  |  |  | J1A |  |  |  |  | A | J |  | A |
| 9A12 | 騒 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 9A13 | 験 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 9A28 | 騨 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 9C04 | 鰄 |  |  |  |  | J1A |  |  |  |  | C | J |  |  |
| 9C2E | 鰮 |  |  |  |  | J1A |  |  |  |  | C | J |  |  |
| 9C76 | 鱶 |  |  |  |  | J1A |  |  |  |  | C | J |  |  |
| 9D0E | 鴎 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 9D2C | 鴬 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 9D8F | 鶏 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 9DC0 | 鷀 |  | V |  |  |  |  |  |  |  |  |  |  |  |
| 9E78 | 鹸 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |
| 9EB9 | 麹 |  | N |  |  | J1A |  |  |  |  | A | J |  |  |
| 9EBA | 麺 |  |  |  |  | J1A |  |  |  |  | A | J |  |  |

The above table contains 172 Hanzi characters outside CGP R0, including Hong Kong local characters that HKIRC submit to CDNC from 2013 to 2015 but not accepted by CDNC, NHCU characters fall in the range of MSR, IICORE characters overlap with MSR. CGP and CDNC held the joint meeting and invited experts from China mainland, Taiwan and Hong Kong to review all these 172 characters and output a variant mapping review document as Appendix G. Based on this variant mapping review document, CGP imported necessary variant characters and adjust related variant mappings.

In Haikou, May 2016, CGP & CDNC joint meeting reviewed 7 Hanzi characters which are non-IICORE but included in dotAsia IDN Table.



The variant mappings of 7 characters were reviewed [on](http://cn.bing.com/dict/search?q=on&FORM=BDVSP6&mkt=zh-cn) the [following](http://cn.bing.com/dict/search?q=following&FORM=BDVSP6&mkt=zh-cn) [basis](http://cn.bing.com/dict/search?q=basis&FORM=BDVSP6&mkt=zh-cn):

* Decouple the variant mappings cross-plane (Panle0 and Plane2)
* Merge correlative variant characters into a union set
* Reset the preferred-simp and preferred-trad for 7 characters and their variants

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| code point | Char | Table | Simp | Trad | other variatn |  | Simp | Trad | other variatn |
| 39DB | 㧛 | dotAsia | 擥(64E5) | 揽(63FD) 攬(652C) |  | >> | 㧛(39DB) | 擥(64E5) | 㩜(3A5C) 揽(63FD) 攬(652C) |
| 3BA3 | 㮣 | dotAsia | 㮣(3BA3) | 㮣(3BA3) | 槩(69E9) | 㮣(3BA3) | 㮣(3BA3) | 槩(69E9) |
| 43D3 | 䏓 | dotAsia | 䏓(43D3) | 䏓(43D3) | 朊(670A) | 䏓(43D3) | 䏓(43D3) | 朊(670A) |
| 4443 | 䑃 | dotAsia | 䑃(4443) | 䑃(4443) | 朦(6726) | 䑃(4443) | 䑃(4443) | 朦(6726) |
| 4882 | 䢂 | dotAsia | 䢂(4882) | U+282E2 |  | 䢂(4882) | 䢂(4882) |  |
| 4C9D | 䲝 | dotAsia | 䲝(4C9D) | 䱽(4C7D) |  | 䲝(4C9D) | 䱽(4C7D) | 鲳(9CB3) 鯧(9BE7) |
| 4C9E | 䲞 | dotAsia | 䲞(4C9E) | U+29D98 |  | 䲞(4C9E) | 䲞(4C9E) |  |
| 652C | 攬 | CDNC | 揽(63FD) | 攬(652C) | 㩜(3A5C) 擥(64E5) | 揽(63FD) | 攬(652C) | 㧛(39DB) 擥(64E5) 㩜(3A5C) |
| 64E5 | 擥 | CDNC | 擥(64E5) | 擥(64E5) | 㩜(3A5C) 揽(63FD) 攬(652C) | 擥(64E5) | 擥(64E5) | 㧛(39DB) 㩜(3A5C) 揽(63FD) 攬(652C) |
| 63FD | 揽 | CDNC | 揽(63FD) | 攬(652C) | 㩜(3A5C) 擥(64E5) | 揽(63FD) | 攬(652C) | 㧛(39DB) 㩜(3A5C) 擥(64E5) |
| 3A5C | 㩜 | CDNC | 㩜(3A5C) | 㩜(3A5C) | 揽(63FD) 擥(64E5) 攬(652C) | 㩜(3A5C) | 㩜(3A5C) | 㧛(39DB) 擥(64E5) 揽(63FD) 攬(652C) |
| 69E9 | 槩 | CDNC | 槩(69E9) | 槩(69E9) |  | 槩(69E9) | 槩(69E9) | 㮣(3BA3) |
| 670A | 朊 | CDNC | 朊(670A) | 朊(670A) |  | 朊(670A) | 朊(670A) | 䏓(43D3) |
| 6726 | 朦 | CDNC | 朦(6726) | 朦(6726) |  | 朦(6726) | 朦(6726) | 䑃(4443) |
| 9CB3 | 鲳 | CDNC | 鲳(9CB3) | 鯧(9BE7) | 䱽(4C7D) | 鲳(9CB3) | 鯧(9BE7) | 䲝(4C9D) 䱽(4C7D) |
| 9BE7 | 鯧 | CDNC | 鲳(9CB3) | 鯧(9BE7) | 䱽(4C7D) | 鲳(9CB3) | 鯧(9BE7) | 䲝(4C9D) 䱽(4C7D) |
| 4C7D | 䱽 | 172 | 䲝(4C9D) | 䱽(4C7D) | 鲳(9CB3) 鯧(9BE7) | 䲝(4C9D) | 䱽(4C7D) | 鲳(9CB3) 鯧(9BE7) |
| 282E2 |  | dotAsia | 䢂(4882) | U+282E2 |  | 282E2 | 282E2 |  |
| 29D98 |  | dotAsia | 䲞(4C9E) | U+29D98 |  | 29D98 | 29D98 |  |

After the above two reviews, the variant mapping table corresponding to CGP R3 was generated as **Appendix H**. Appendix H complies with CDNC rules. The variant mappings in the final XML document will be somewhat different from Appendix H, not only due to the sub-type tags in Section 4.1, moreover, some new sub-type tags will be created to avoid overproducing allocatable labels, which will be illustrated in Section 4.6.

## 4.4 Variant inconsistency between .asia and CGP

In CGP R3 and its variant table, there are 69 characters whose variant mappings are different from the counterparts in .asia IDN Table. This inconsistency reflects the [regional](http://cn.bing.com/dict/search?q=Regional&FORM=BDVSP6&mkt=zh-cn) [cultural](http://cn.bing.com/dict/search?q=Cultural&FORM=BDVSP6&mkt=zh-cn) individuality and internet application difference on SLD registration in the early days. CGP and Edmon CHUNG discussed the issue and agreed that the DotAsia table has been created as experimental for the HK characters, the intent has always been to merge and make consistent with CGP table once it is integrated for root zone and gTLD purpose. The 69 characters and the variant mappings are as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 3A18 | 㨘 | 擤(64E4) | 㨘(3A18) | 㨘(3A18)擤(64E4) |
| 3A52 | 㩒 | 擒(64D2) | 㩒(3A52) | 㩒(3A52)擒(64D2) |
| 4882 | 䢂 | 䢂(4882) | 䢂(4882) | 䢂(4882) |
| 4C7D | 䱽 | 鲳(9CB3) | 鯧(9BE7) | 䱽(4C7D)鯧(9BE7)鲳(9CB3) |
| 4C9D | 䲝 | 䲝(4C9D) | 䲝(4C9D) | 䲝(4C9D) |
| 4C9E | 䲞 | 䲞(4C9E) | 䲞(4C9E) | 䲞(4C9E) |
| 5047 | 假 | 假(5047) | 假(5047) | 假(5047) |
| 51C0 | 净 | 净(51C0) | 淨(6DE8) | 净(51C0)凈(51C8)浄(6D44)淨(6DE8) |
| 51C8 | 凈 | 净(51C0) | 凈(51C8) | 净(51C0)凈(51C8)浄(6D44)淨(6DE8) |
| 52A4 | 劤 | 劤(52A4) | 劤(52A4) | 劤(52A4) |
| 52B2 | 劲 | 劲(52B2) | 勁(52C1) | 劲(52B2)勁(52C1) |
| 52C1 | 勁 | 劲(52B2) | 勁(52C1) | 劲(52B2)勁(52C1) |
| 53DA | 叚 | 叚(53DA) | 叚(53DA) | 叚(53DA)段(6BB5) |
| 58DC | 壜 | 壜(58DC) | 壜(58DC) | 坛(575B)埮(57EE)墰(58B0)壇(58C7)壜(58DC)罈(7F48)罎(7F4E) |
| 5B0E | 嬎 | 嬎(5B0E) | 嬎(5B0E) | 嬎(5B0E) |
| 5B14 | 嬔 | 嬔(5B14) | 嬔(5B14) | 嬔(5B14) |
| 5BD7 | 寗 | 寗(5BD7) | 寗(5BD7) | 宁(5B81)寍(5BCD)寕(5BD5)寗(5BD7)寜(5BDC)寧(5BE7)甯(752F) |
| 60E3 | 惣 | 总(603B) | 總(7E3D) | 总(603B)惣(60E3)捴(6374)搃(6403)摠(6460)総(7DCF)縂(7E02)總(7E3D) |
| 617D | 慽 | 慽(617D) | 慽(617D) | 慼(617C)慽(617D) |
| 621A | 戚 | 戚(621A) | 戚(621A) | 戚(621A) |
| 6335 | 挵 | 弄(5F04) | 挵(6335) | 弄(5F04)挵(6335)衖(8856) |
| 637F | 捿 | 栖(6816) | 棲(68F2) | 捿(637F)栖(6816)棲(68F2) |
| 63D4 | 揔 | 揔(63D4) | 揔(63D4) | 揔(63D4) |
| 64E1 | 擡 | 抬(62AC) | 擡(64E1) | 抬(62AC)擡(64E1) |
| 64E5 | 擥 | 擥(64E5) | 擥(64E5) | 㧛(39DB)擥(64E5)揽(63FD)攬(652C) |
| 656D | 敭 | 扬(626C) | 敭(656D) | 扬(626C)揚(63DA)敭(656D)颺(98BA)飏(98CF) |
| 6667 | 晧 | 皓(7693) | 皓(7693) | 晧(6667)暠(66A0)皓(7693)皜(769C) |
| 67A3 | 枣 | 枣(67A3) | 棗(68D7) | 枣(67A3)栆(6806)棗(68D7) |
| 681E | 栞 | 刊(520A) | 栞(681E) | 刊(520A)刋(520B)栞(681E) |
| 68C5 | 棅 | 柄(67C4) | 柄(67C4) | 柄(67C4)棅(68C5) |
| 6900 | 椀 | 椀(6900) | 碗(7897) | 椀(6900)盌(76CC)碗(7897)鋺(92FA) |
| 6A53 | 橓 | 蕣(8563) | 蕣(8563) | 橓(6A53)蕣(8563) |
| 6D44 | 浄 | 净(51C0) | 淨(6DE8) | 净(51C0)凈(51C8)浄(6D44)淨(6DE8) |
| 6DE8 | 淨 | 净(51C0) | 淨(6DE8) | 净(51C0)凈(51C8)浄(6D44)淨(6DE8) |
| 6EDD | 滝 | 泷(6CF7) | 瀧(7027) | 泷(6CF7)滝(6EDD)瀧(7027) |
| 701E | 瀞 | 瀞(701E) | 瀞(701E) | 瀞(701E) |
| 7200 | 爀 | 赫(8D6B) | 赫(8D6B) | 爀(7200)赫(8D6B) |
| 734F | 獏 | 貘(8C98) | 貘(8C98) | 獏(734F)貘(8C98) |
| 73E1 | 珡 | 琴(7434) | 琴(7434) | 珡(73E1)琴(7434) |
| 73E4 | 珤 | 宝(5B9D) | 寶(5BF6) | 宝(5B9D)寳(5BF3)寶(5BF6)珤(73E4) |
| 7460 | 瑠 | 琉(7409) | 瑠(7460) | 琉(7409)瑠(7460)璢(74A2) |
| 74C8 | 瓈 | 璃(7483) | 琍(740D) | 琍(740D)璃(7483)瓈(74C8) |
| 757A | 畺 | 疆(7586) | 疆(7586) | 畺(757A)疆(7586) |
| 764E | 癎 | 痫(75EB) | 癇(7647) | 痫(75EB)癇(7647)癎(764E) |
| 767A | 発 | 发(53D1) | 發(767C) | 发(53D1)彂(5F42)発(767A)發(767C)髪(9AEA)髮(9AEE) |
| 76CC | 盌 | 碗(7897) | 碗(7897) | 椀(6900)盌(76CC)碗(7897)鋺(92FA) |
| 7AC3 | 竃 | 灶(7076) | 竈(7AC8) | 灶(7076)竃(7AC3)竈(7AC8) |
| 7B6F | 筯 | 箸(7BB8) | 箸(7BB8) | 筯(7B6F)箸(7BB8) |
| 7B92 | 箒 | 帚(5E1A) | 帚(5E1A) | 帚(5E1A)箒(7B92)菷(83F7) |
| 7C83 | 粃 | 秕(79D5) | 秕(79D5) | 秕(79D5)粃(7C83) |
| 7DDC | 緜 | 绵(7EF5) | 綿(7DBF) | 綿(7DBF)緜(7DDC)绵(7EF5) |
| 8117 | 脗 | 吻(543B) | 吻(543B) | 吻(543B)呅(5445)呡(5461)脗(8117) |
| 840C | 萌 | 萌(840C) | 萌(840C) | 萌(840C)蕄(8544) |
| 8420 | 萠 | 萠(8420) | 萠(8420) | 萠(8420) |
| 84DA | 蓚 | 蓨(84E8) | 蓨(84E8) | 蓚(84DA)蓨(84E8) |
| 8544 | 蕄 | 萌(840C) | 蕄(8544) | 萌(840C)蕄(8544) |
| 8597 | 薗 | 园(56ED) | 園(5712) | 园(56ED)園(5712)薗(8597) |
| 89A9 | 覩 | 睹(7779) | 睹(7779) | 睹(7779)覩(89A9) |
| 8EE2 | 転 | 转(8F6C) | 轉(8F49) | 転(8EE2)轉(8F49)转(8F6C) |
| 8FBA | 辺 | 边(8FB9) | 邊(908A) | 边(8FB9)辺(8FBA)邉(9089)邊(908A) |
| 9244 | 鉄 | 铁(94C1) | 鐵(9435) | 鉄(9244)銕(9295)鐡(9421)鐵(9435)铁(94C1) |
| 932C | 錬 | 炼(70BC) | 煉(7149) | 炼(70BC)煉(7149)錬(932C)鍊(934A) |
| 94C1 | 铁 | 铁(94C1) | 鐵(9435) | 鉄(9244)銕(9295)鐡(9421)鐵(9435)铁(94C1) |
| 9771 | 靱 | 韧(97E7) | 韌(97CC) | 肕(8095)靭(976D)靱(9771)韌(97CC)韧(97E7) |
| 98C7 | 飇 | 飙(98D9) | 飆(98C6) | 飆(98C6)飇(98C7)飈(98C8)飙(98D9)飚(98DA) |
| 994D | 饍 | 膳(81B3) | 膳(81B3) | 膳(81B3)饍(994D) |
| 99C5 | 駅 | 驿(9A7F) | 驛(9A5B) | 駅(99C5)驛(9A5B)驿(9A7F) |
| 9D44 | 鵄 | 鸱(9E31) | 鴟(9D1F) | 鴟(9D1F)鵄(9D44)鸱(9E31) |
| 9F62 | 齢 | 龄(9F84) | 齡(9F61) | 齡(9F61)齢(9F62)龄(9F84) |

## 4.5 Coordination on unacceptable variant mappings for KGP

Another focus issue is that not all variant mappings will get agreed by three parties, some variant mappings are totally UNACCEPTABLE to one party. Some characters in a CGP variant group have the same pronunciations and meanings, but have different meanings in other language environments. For example, in Korean, (U+673机) means [desk, small table] and (U6A5F機) means [machine], but both mean [machine] in Chinese. In the CJK coordination meeting, March 2016, KGP raised 258 Hanja/Hanzi variant group [Appendix X] whose variant mappings in CGP LGR are UNACCEPTABLE. As the requesting party for the vast majority of Chinese character variants, CGP has the obligation to take the initiative to move the work forward. CGP has made statistics and analysis of all disputed 258 variant groups, including the number of registered labels containing disputed character under .CN/.TW/.HK/.网址, and the semantics of all these registered labels.

CGP would take different coordination approaches for different situation:

* Decouple the variant mapping and keep the characters in repertoire independently.
* Remove the characters from CGP repertoire
* Persuade other parties into accepting CGP variant mapping

In August 2016, CGP proposed a list of 141 variant groups to KGP. CGP insists these 141 variant groups go with CGP rules, while leaves the others to KGP. CJK will hold 3rd joint meeting in Taipei at the end of September and hope three parties could reach a consensus on the disposition of 258 variant groups.

## 4.6 Variant Mappings that overproduce allocatable labels

Traditionally in CDNC rule, the variant mappings and WLE rules are designed with the assumption that given any valid input label, there would be at most three resulting allocatable labels -- the original label, an all-simplified label, and an all-traditional label.

In Appendix H, for all 19746 characters, 10 have 2 simplified variants, 175 have 3 traditional variants, 9 have 4 traditional variants and 2 have four traditional variants, which means, all together 196 characters have multiple allocatable variant mappings.

These 196 code points with multiple allocatable variant mappings will generate multiple all-simplified labels or all-traditional labels, which will violate CDNC constraint and bring over-production issue at the root zone level. In SLD practice, CDNC members designed extra human interaction mechanism to avoid the dilemma, to enable the applicants to SELECT at most one all-simplified and at most one all-tradtional from the multiple alternatives. Once [the](http://cn.bing.com/dict/search?q=the&FORM=BDVSP6&mkt=zh-cn) [selection](http://cn.bing.com/dict/search?q=selection&FORM=BDVSP6&mkt=zh-cn) is complete, all the other allocatable labels will be reserved. Unlike the blocked labels, these reserved allocatable labels might be activated at the request of applicant.

But at the Root Zone level, without any “human interaction” mechanism, multiple preferred mappings would lead to overproduction of variant labels with an "allocatable" status. An example would be “Taiwan Railway” 台(53F0)鉄(9244).

<char cp="53F0" tag="sc:Hani" ref="0 100 101 102 103 104" >

<var cp="53F0" type="r-both" comment="identity" />

<var cp="6AAF" type="trad" />

      <var cp="7C49" type="blocked" />

      <var cp="81FA" type="trad" />

      <var cp="98B1" type="trad" />

    </char>

<char cp="9244" tag="sc:Hani" ref="0 101 102 104" >

<var cp="9244" type="r-neither" comment="identity,reflexive" />

<var cp="9295" type="blocked" />

<var cp="9421" type="blocked" />

<var cp="9435" type="trad" />

<var cp="94C1" type="simp" />

</char>

An input of 台(53F0)鉄(9244) results in 6 allocatable variant labels:

Original: 台(53F0)鉄(9244)

All-Simplified: 台(53F0)铁(94C1)

All-Traditional: 台(53F0)鐵(9435), 檯(6AAF)鐵(9435), 臺(81FA)鐵(9435), 颱(98B1)鐵(9435)

Some of the above "allocatable" labels might be unnecessary from a semantic standpoint.

According to IP’s suggestion in feedback document for CGP LGR proposal 20160613, under the conservatism principle, LGRs should strive to minimize allocatable variants, and most of these cases can be fixed by not having multiple simp/trad mappings. It may be an acceptable trade-off to eliminate the multiple mappings, and let applicants who need a specific all-simplified or all-traditional variant label apply for just the specific label.

To check the feasibility of the above mechanism, CGP take 台(53F0)鉄(9244) as example:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| A | B | | C | | D | | E |
| 台(53F0) | 檯(6AAF) | | 籉(7C49) | | 臺(81FA) | | 颱(98B1) |
| Allocatable Mappings | | Original Char | | Simplified mappings | | Traditional mappings | |
| A; A, ABDE | | 台(53F0) | | 台(53F0) | | 台(53F0)檯(6AAF)  臺(81FA)颱(98B1) | |
| B; A, B | | 檯(6AAF) | | 台(53F0) | | 檯(6AAF) | |
| C; A, C | | 籉(7C49) | | 台(53F0) | | 籉(7C49) | |
| D; A, D | | 臺(81FA) | | 台(53F0) | | 臺(81FA) | |
| E; A, E | | 颱(98B1) | | 台(53F0) | | 颱(98B1) | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| L | M | | N | | O | | P |
| 鉄(9244) | 銕(9295) | | 鐡(9421) | | 鐵(9435) | | 铁(94C1) |
| Allocatable Mappings | | Original Char | | Simplified mappings | | Traditional mappings | |
| L; P, O | | 鉄(9244) | | 铁(94C1) | | 鐵(9435) | |
| M; P, M | | 銕(9295) | | 铁(94C1) | | 銕(9295) | |
| N; P, O | | 鐡(9421) | | 铁(94C1) | | 鐵(9435) | |
| O; P, O | | 鐵(9435) | | 铁(94C1) | | 鐵(9435) | |
| P; P, O | | 铁(94C1) | | 铁(94C1) | | 鐵(9435) | |

It might be safe to remove reflexive “A” from the multiple mappings (change the reflexive sub-type of “A” into some new one), because no matter if the reflexive variant exists, the applied-for label will be generated automatically and contain the original “A”.

|  |  |
| --- | --- |
| Old Mappings | New Mappings |
| A;A,ABDE | A;A,BDE |

But in the real registration scenarios, this removal could end up causing serious problems.

With the old mappings (A; A, ABDE) and (L; P, O), the applicant inputs “台铁/AL” and gets:

* one original label: “AL”
* one all-simplified label: “AP”
* 4 all-traditional labels: “AO” “BO” “DO” “EO”

With the new mappings (A; A, BDE) and (L; P, O), the applicant inputs “AV” and only get:

* one original label: “AL”
* one all-simplified label: “AP”
* 4 all-traditional labels: “BO” “DO” “EO”

For an applicant who expects all-simplified “AZ”台(53F0)鉄(9244) and all-traditional “AO”台(53F0) 鐵(9435), the removal of reflexive “A” from multiple mappings will make it impossible to generate “AO”.

If applicant apply for the specific label of “AO”, as IP suggested, with (A; A, BDE) and (O; P, O), he/she will only get:

* one original label: “AO”
* one all-simplified label: “AP”
* 4 all-traditional labels: “BO” “DO” “EO”

In this way, the applicant gets all-traditional “AO”, while in exchange, all-simplified “AZ” becomes [unacquirable](http://cn.bing.com/dict/search?q=unacquirable&FORM=BDVSP6&mkt=zh-cn).

Actually, for any given label “XY”, when Y’s variant mappings go with (Y; Ys, Yt) and Y≠Ys≠Yt (there are 423 such characters), removing reflexive variant “X” from multiple variant mappings is dangerous, not to mention removing non-reflexive variant. It is an aggressive and risky solution to limit the number of allocatable labels by eliminate the multiple mappings, because the LGR can hardly predict which all-simplified or all-traditional label is exactly wanted by different applicant in different context.

In practice, CDNC members only delegate the applied label, one preferred SC label and one preferred TC label, no more than THREE labels to the applicants, and reserve the others. Some supplementary measures are introduced to reduce the number of allocatable labels and related computational complexity, like human interaction process (menu of choices for user input) or post-delegation activation process (activate reserved labels as requested). While in current XML-format transforming regulations, the BLOCKED labels can never be allocated to anyone under any circumstances. If CGP WLE is supposed to limit the number of allocatable labels, some similar CDNC design should be taken into account, which correspondingly will change the way LGR works.

Here is a temporary solution CGP would like to propose.

If IP insist that eliminating multiple mappings is necessary and essential, CGP will try to create 6 new sub-types:

|  |  |  |
| --- | --- | --- |
| Sub-Type | Type | Comment |
| “r-simp-m” | Blocked | In multiple simplified mappings, change reflexive variant char from “r-simp” into “r-simp-m” |
| “simp-m” | Blocked | In multiple simplified mappings, change non-reflexive char from “simp” into “simp-m” until one is left.  In practice, the char with the smallest hex-code will be left. |
| “r-trad-m” | Blocked | In multiple traditional mappings, change reflexive variant char from “r-trad” into “r-trad-m” |
| “trad-m” | Blocked | In multiple traditional mappings, change non-reflexive char from “trad” into “trad-m” until one is left.  In practice, the char with the smallest hex-code will be left. |
| “r-both-m” | Blocked | r-smip-m char and r-trad-m char is the same |
| “both-m” | Blocked | smip-m char and trad-m char is the same |

By introducing the above 6 sub-type, all characters will have only one simplified variant and one traditional variant, then the new LGR could ultimately limit the number of allocatable labels under THREE, but on the other end, it could also prevent an applicant from getting preferred all-simplified or preferred all-traditional labels very likely.

However, as a corresponding **compensatory measure** to the above sub-types, LGR is needed to be executed for MULTIPLE times (at most THREE times) to allow the applicant input different variant forms of applied-for labels, by merging the outputs (at most FIVE labels), to acquire his preferred allocatable labels. CGP noticed that JGP has the similar concern and requirement for MULTIPLE LGR execution. Since Japanese has no variants (each character is independent), theoretically, it is impossible to define variant type (like new/old form) and processing rule for imported variants from CGP/KGP's LGR. Any attempt to define a “blocked” sub-type in JGP LGR might block the real wanted whole label with a great possibility.

Over all, in CGP LGR 20160913 (Appendix I), CGP decided to introduce 6 new subtypes (“simp-m”, “trad-m”, “r-simp-m”, “r-trad-m”, “r-both-m”) and, correspondingly, add new rules for these sub-types. But before the compensation proposal were accepted, the new sub-types will not be activated, and no variant chars will be tagged with them.

# Whole Label Evaluation Rules

Following XML-format transforming regulations (Representing Label Generation Rulesets using XML, <https://datatracker.ietf.org/doc/draft-davies-idntables/>), CGP generates its own XML table of CGP repertoire and variant mappings, and marks every variant code point with the following tags:

”r-simp”, “r-trad”, ”r-both”

“simp”, “trad”, “both”

“r-simp-m”, “r-trad-m”, “r-both-m”

“simp-m”, “trad-m”, “both-m”

“r-neither”, “blocked”

Consistent with CDNC’s “TC-SC Equivalence” rule, delegating the applied label, one preferred SC label and one preferred TC label to the same applicant, CGP WLE is designed as follows:

<rules>

<!--Action elements - order defines precedence-->

<action disp="invalid" match="leading-combining-mark" comment="labels with leading combining marks are invalid" />

<action disp="blocked" any-variant="blocked" comment="default action for blocked variant"/>

<action disp="allocatable" only-variants="simp r-simp both r-both" comment="simplified label" />

<action disp="allocatable" only-variants="trad r-trad both r-both" comment="traditional label"/>

<action disp="allocatable" only-variants="r-simp r-trad r-both r-neither" comment="original label"/>

<action disp="blocked" only-variants="simp simp-m r-simp r-simp-m both both-m r-both r-both-m" comment="multiple simplified label" />

<action disp="blocked" only-variants="trad trad-m r-trad r-trad-m both both-m r-both r-both-m" comment="multiple traditional label"/>

<action disp="blocked" any-variant="simp trad both r-simp r-trad r-both simp-m trad-m both-m r-simp-m r-trad-m r-both-m r-neither" comment="block any other mixed labels" />

<action disp="allocatable" comment="catch-all" />

</rules>

As mentioned in Section 4.6, new rules were created for 6 new subtypes (“simp-m”, “trad-m”, “r-simp-m”, “r-trad-m”, “r-both-m”) , but they don’t really work because no variant chars are tagged with these subtype in the current Variant Mappings (Appendix I CGP Variant Mappings 20160923).

# References

The Unicode Standard 1.1

The Unicode Standard 2.0

The Unicode Standard 2.1

The Unicode Standard 3.0

The Unicode Standard 3.1

The Unicode Standard 3.2

The Unicode Standard 4.0

The Unicode Standard 4.1

The Unicode Standard 5.0

The Unicode Standard 5.1

The Unicode Standard 5.2

The Unicode Standard 6.0

The Unicode Standard 6.1

The Unicode Standard 6.2

The Unicode Standard 6.3

CDNC IDN Table <http://www.cdnc.org/gb/research/file/unicode.txt>

International Ideographs Core <http://appsrv.cse.cuhk.edu.hk/~irg/irg/IICore/IRGN1067R2_IICore22_MappingTable.txt>

China's State Council Normalized Hanzi List for Common Use <http://www.gov.cn/zwgk/2013-08/19/content_2469793.htm>

DotAsia ZH IDN Table <http://www.iana.org/domains/idn-tables/tables/asia_zh_1.1.txt>

## 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>
* K. Davies, A. Freytag, Representing Label Gneration Rulesets using XML, <https://datatracker.ietf.org/doc/draft-davies-idntables/>

## ICANN Related Documents

* ICANN.Guidelines for the Implementation of Internationalised Domain Names (2003). <http://www.icann.org/en/general/idn-guidelines-20jun03.htm>
* ICANN. New gTLD draft Applicant Guidebook. 2011, <http://www.icann.org/en/topics/new-gtlds/rfp-clean-19sep11-en.pdf>
* ICANN, Chinese Case Study Team Report, Report on Chinese Variants in Internationalized Top-Level Domains, 2011, < <https://archive.icann.org/en/topics/new-gtlds/chinese-vip-issues-report-03oct11-en.pdf>>

# Appendix A：CGP Repertoire

The current membership of the Chinese Script Generation Panel (CGP) includes the following (in alphabetical order).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No.** | **Name** | **Organization** | **Country/Region** | **Language Expertise** |
| **1** | Chao QI | CNNIC | China | Chinese |
| **2** | Chris DILLON | University College London | UK | Chinese/Japanese/Korean |
| **3** | Connie Hon | IP Mirror | Singapore | Chinese |
| **4** | Di MA | ZDNS | China | Chinese |
| **5** | Guoying LI | Beijing Normal University | China | Chinese |
| **6** | Holmes LEONG | MONIC | Macao | Chinese |
| **7** | James SENG | 21ViaNet Group Limited | Malaysia | Chinese |
| **8** | Jean-Jacques Subrenat | ATLAC ICANN | France | French, English, Chinese, Japanese. |
| **9** | Jenifer CHUNG | Dot Asia | USA/Hongkong | Chinese |
| **10** | Jiagui XIE | CONAC | China | Chinese |
| **11** | Jonathan SHEA | HKIRC | Hong Kong | Chinese |
| **12** | Joseph YEE | Afilias | Canada | Simplified Chinese, Traditional Chinese, (Familiar with Japanese) |
| **13** | Kenny HUANG | TWNIC | Taiwan | Chinese |
| **14** | Linlin ZHOU | CNNIC | China | Chinese |
|  | Lu QIN | Hong Kong Polytechnic University | Hong Kong | Chinese |
| **15** | Nai-Wen HSU | TWNIC | Taiwan | Chinese |
| **16** | Ryan TAN | SGNIC | Singapore | Chinese |
| **17** | Shutian CUI | Ministry of Industry and Information Technology | China | Chinese |
| **18** | Wei WANG | CNNIC | China | Chinese |
| **19** | Xiaodong LEE | CNNIC | China | Chinese |
| **20** | Yuxiao LI | Beijing University of Posts and Telecommunications | China | Chinese |
| **21** | Zheng WANG |  | China | Chinese |
| **22** | Zhiwei YAN | CNNIC | China | Chinese |
| **23** | Zhoucai ZHANG | UniHan Digital Tech., Ltd. | China | Chinese mainly |

The Generation Panel (GP) for the Chinese Script LGR gathers experts from a variety of backgrounds (bringing varied linguistic and technical perspectives), including those who are national and regional policy makers, members from the technical community directly working with the DNS (e.g. registries and registrars), security, academia (technical and linguistic), members of community based organizations, and members with experience of local language studying.

Geographically, the GP for the Chinese script has members from across the relevant regions, including East Asia and Southeast Asia. There are also experts from non-Chinese-speaking regions equipped with profound knowledge in oriental languages as well as culture. The members belong to seven different countries/regions from these areas. The members and their expert backgrounds are listed as follows:

**Wei Wang** is the chair of CGP. He is the deputy chief engineer of CNNIC and former deputy director of CNNIC. As the co-secretary of CDNC, he worked as a member of ICANN’s Chinese VIP team. His expert background in Chinese domain name registration and management helps to control the overall progress of the CGP, and to propose and evaluate the key schemes and policies.

**Kenny Huang** is the co-chair of the CGP, in charge of the coordination between C, J and K. He is on the APNIC executive council, the DotAsia advisory council and is a TWNIC board director and chair of the International Affair Committee. He worked as the APNIC policy SIG chair from 2002 to 2007 and on ICANN’s ASO Address Council from 2001 to 2011. He is the co-author of IETF RFC 3743. He is an IDN expert and helps to coordinate with the JGP and KGP to regulate Chinese variant mapping rules.

**Xiaodong Lee** is a policy expert. He is the CEO and CTO of CNNIC. He acts as the former vice president of ICANN and is also the organizer of several international and domestic technology standards in the fields of domain names and email. He is the co-author of RFC 4713 and is the CGP’s policy expert.

**Jean-Jacques SUBRENAT is a representative from ATLAC, and also from the European community.** He is the member of the "NTIA IANA Functions' Stewardship Transition Coordination Group"; member of ALAC (2010-12, again 2012-14); former member of ICANN Board (2007-10); former diplomat (1971-2005) and retired Ambassador (1998-2005). His working languages include French, English, Chinese and Japanese.

**Chris Dillon is a linguistic expert.** He was a member of ICANN’s Chinese VIP team. He was a member of the joint ccNSO/gNSO IDN working group, is Co-Chair of the GNSO’s Translation and Transliteration PDP WG and provides linguistic advices to the CGP. As a non-native Chinese linguist, he provides valuable professional opinions from the European community.

**Guoying LI** is a linguistic expert. He is the Vice-Dean of the School of Chinese Language and Literature of Beijing Normal University. He is also the director of the Research Center of Classification and Standardization of Chinese Characters, supported by Beijing Normal University and the Language and Information Division, Ministry of Education of China.

**Joe Zhang** **is a Chinese script processing expert**. From 1989 to 2003, he was the key developer of ISO/IEC 10646, chaired the CJK group as CJK JRG, and is an IRG rapporteur and contributing editor. As the CEO of UniHan Digital Technology, he acts as the invited researcher of the Language Application Institute under the Ministry Education of China. His background of Chinese language and Chinese culture provides support for variants set regulation.

**Lu QIN is a linguistic expert** Professor. She has successfully helped to make structured encoding of Chinese character components that lead to a much faster encoding process for Chinese characters. She spearheaded the standardization of the Hong Kong Supplementary Character Set, the first and only commonly adopted character set for Hong Kong. She is the Rapporteur of the ISO/IEC JTC1/SC2/WG2/IRG.

**Shutian CUI** is a government representative. She serves as the Division Director at the Telecommunication Administration Bureau of Ministry of Industry and Information Technology (MIIT), PRC, involving namely registry and registrar policy making, supervision on the DNS operation, protection of the public interest and the user's information, development of the Chinese TLDs, coordination of the IP address designation and allocation, etc.

**Yuxiao LI** is a [legal expert](http://dict.cn/legal%20expert). He is the Dean at Institute of Cyber Governance and Law in Beijing University of Posts and Telecommunications (BUPT), China. He has rich experience on Internet governance and law research.

**Chao QI** is a registry expert. He is a R&D engineer of CNNIC, takes responsibility for SRS, RDDS and DNS systems for CC TLD and New gTLD and has practical experience of Chinese variant issues in the registration for Chinese domain name.

**Jonathan SHEA** is a registry expert. He is the CEO of HKIRC and HKDNR. He was a member of ICANN’s Chinese VIP team. From 2012, he was a member of the Joint ccNSO/gNSO IDN Working Group. He is the representative of registries/registrar in Hong Kong and also provides advice for the regulation of variants set and rules.

**Jiagui XIE** is a DNS expert. He is the director of the new gTLD Lab in CONAC (the 政务 and 公益 new gTLD Registry). He has rich experience in DNS/gTLD/ccTLD/IDN and EPP technologies.

**Linlin ZHOU** is an IDNA expert. She is the author of several Whois-related IETF drafts with working-group status.

**Nai-Wen HSU** is a DNS/IDNA/Unicode expert. He is the technical department director at TWNIC and co-author of RFC4713. He helps to evaluate the variant mapping rules.

**Ryan Tan is a registry/registrar expert.** Since 2000 he has been involved in IDNs of various languages in both technical and policy aspects. In 2005 he pioneered the Chinese and Tamil IDN testbed for SGNIC which laid the ground work for the eventual launch of Chinese.com.sg and Chinese.sg domain names in 2009 and Chinese.新加坡 and Tamil.சிங்கப்பூர் in 2011.

**Zhiwei YAN** is a DNS expert. He is in charge of the DNS and IPv6 researches in CNNIC and he is also the invited professor of Waseda University (Japan) since 2013.

**Zheng WANG is a representative for China mainland community.** He used to be the director of joint labs in CONAC and is also the architect of CONAC’s registry system. He is a member of ICANN’s Internationalized Registration Data Working Group.

**Di MA** is a registrar expert. He is the laboratory director of the Internet Domain Name System Beijing Engineering Research Centre (ZDNS), the ICANN accredited registrar and New gTLD back-end service provider hosting over 20 new gTLDs including more than 10 Chinese gTLDs.

**Joseph YEE** is a IDNA/Unicode expert. He works for Afilias and is also an expert in Japanese language and provides advice for Chinese characters usage in the Japanese language.

**Connie HON** is a representative for Singapore community. She works in Business Development and Strategy with IP Mirror Pte Limited.

**Holmes LEONG** is a representative for Macao community. He is the Chief Operations Officer at HNET Asia Limited.

**Jennifer CHUNG** is the representative for Hong Kong community and USA community. She is the current Policy and Organisational Relations lead for DotAsia Organisation and based in the US.  She is a member of the Translation and Transliteration of Contact Information PDP working group within the GNSO.‏

**James SENG** is a representative for Malaysia community. James also participates actively in several standard organizations (such as ISO/IEC JTC1 and IETF) and also served on the board/committee of several Internet organizations.

# Appendix B: CGP Repertoire

# Appendix C: JGP Repertoire

# Appendix D: KGP Repertoire

# Appendix E: KGP variant Mappings

# Appendix F: KGP Unacceptable 258 Variant Mappings

# Appendix G: CGP Review on 172 Hanzi Characters

# Appendix H: CGP Variant Mappings

# Appendix I: CGP LGR