**Proposal for Chinese Script Root Zone LGR**

[0 General Information 3](#_Toc488359228)

[1 Script and Languages Covered 4](#_Toc488359229)

[1.1 Overview 4](#_Toc488359230)

[1.2 Countries with Significant Usage of the Chinese Script 5](#_Toc488359231)

[1.3 Target Script: Hani 6](#_Toc488359232)

[1.4 Principal Languages using the Script 7](#_Toc488359233)

[2 Process for Developing the Proposal 9](#_Toc488359234)

[2.1 Previous work 9](#_Toc488359235)

[2.2 Team Diversity 10](#_Toc488359236)

[2.3 Work Process 10](#_Toc488359237)

[3 Code point repertoire 13](#_Toc488359238)

[3.1 Basic character set 13](#_Toc488359239)

[3.2 Repertoire formation process 13](#_Toc488359240)

[3.2.1 Basic Repertoire 13](#_Toc488359241)

[3.2.2 124 dotAsia characters 14](#_Toc488359242)

[3.2.3 18 TGSCC characters 15](#_Toc488359243)

[3.2.4 43 CJK Coordination characters 16](#_Toc488359244)

[3.3 Attempt to limit the size of the repertoire 18](#_Toc488359245)

[4 Code point variants 24](#_Toc488359246)

[4.1 Variant definition in CGP 24](#_Toc488359247)

[4.2 Variant Mappings formation process 26](#_Toc488359248)

[4.2.1 Basic Variant mappings 26](#_Toc488359249)

[4.2.2 172 TGSCC and IICORE variants review 26](#_Toc488359250)

[4.2.3 69 dotAsia unique variants review 30](#_Toc488359251)

[4.2.4 CJK Variant coordination 32](#_Toc488359252)

[4.3 Multiple Preferred Variant Characters 36](#_Toc488359253)

[5 Whole Label Evaluation Rules 44](#_Toc488359254)

[References 46](#_Toc488359255)

[Internet Drafts and RFCs 46](#_Toc488359256)

[ICANN Related Documents 47](#_Toc488359257)

[Appendix A: CGP Membership 48](#_Toc488359258)

[Appendix B: CGP Repertoire 48](#_Toc488359259)

[Appendix C: JGP Repertoire 48](#_Toc488359260)

[Appendix D: KGP Repertoire 48](#_Toc488359261)

[Appendix E: KGP variant Mappings 48](#_Toc488359262)

[Appendix F: KGP Unacceptable 258 Variant Mappings 48](#_Toc488359263)

[Appendix G: CGP Review on 172 Hanzi Characters 48](#_Toc488359264)

[Appendix H: 445 Variant Mappings Update 48](#_Toc488359265)

[Appendix I: CGP Variant Mappings 48](#_Toc488359266)

[Appendix J: CGP LGR XML 48](#_Toc488359267)

# General Information

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

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

The first version was finished on June 13, 2016 and named after the date 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 on October 12, 2016.

The third version was completed on Jan 18, 2017, trying to address further doubts from the IP in CGP-LGR20160923-Report\_1028.pdf, especially the below:

1. Reduce the size of repertoire based on modern usage analysis of each code point
2. Reduce the number of allocatable labels
3. Explain the inconsistency between CGP and dotAsia IDN Table

CGP and KGP reached consensus on variant mappings in the CK joint meeting, March, Beijing.

CGP participated in the ICANN Helsinki meeting remotely with help from Edmon CHUNG, after that, CGP generated LGR version 4. The [Main](http://cn.bing.com/dict/search?q=Main&FORM=BDVSP6&mkt=zh-cn) [changes](http://cn.bing.com/dict/search?q=changes&FORM=BDVSP6&mkt=zh-cn) [are](http://cn.bing.com/dict/search?q=are&FORM=BDVSP6&mkt=zh-cn) as follows:

1. Section 3.3 explains the CGP’s attempt to reduce the repertoire, and the reason why CGP decided to keep the current size.
2. Section 4.2.3 illustrates the inconsistency between CGP and dotAsia. According to the discussion with Edmon CHUNG, dotAsia would adjust its SLD repertoire and variant mappings to comply with CGP LGR rules once finalized.
3. Section 4.2.4 introduces the coordination outcome between C and K, and the final CJK consensus.
4. Section 4.3 explains the CGP’s attempt to eliminate the multiple mappings and the reason why CGP decided to keep multiple mappings; alternatively, CGP proposed two solutions to limit the number of allocatable labels for further discussion with the IP.

# 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. Hanzi originated from [inscriptions](http://cn.bing.com/dict/search?q=inscriptions&FORM=BDVSP6&mkt=zh-cn) [on](http://cn.bing.com/dict/search?q=on&FORM=BDVSP6&mkt=zh-cn) [bones](http://cn.bing.com/dict/search?q=bones&FORM=BDVSP6&mkt=zh-cn) [or](http://cn.bing.com/dict/search?q=or&FORM=BDVSP6&mkt=zh-cn) [tortoise](http://cn.bing.com/dict/search?q=tortoise&FORM=BDVSP6&mkt=zh-cn) [shells](http://cn.bing.com/dict/search?q=shells&FORM=BDVSP6&mkt=zh-cn) in [the](http://cn.bing.com/dict/search?q=the&FORM=BDVSP6&mkt=zh-cn) [Shang](http://cn.bing.com/dict/search?q=Shang&FORM=BDVSP6&mkt=zh-cn) [Dynasty](http://cn.bing.com/dict/search?q=Dynasty&FORM=BDVSP6&mkt=zh-cn) ([c](http://cn.bing.com/dict/search?q=c&FORM=BDVSP6&mkt=zh-cn). 16[th](http://cn.bing.com/dict/search?q=th&FORM=BDVSP6&mkt=zh-cn)-11[th](http://cn.bing.com/dict/search?q=th&FORM=BDVSP6&mkt=zh-cn) [century](http://cn.bing.com/dict/search?q=century&FORM=BDVSP6&mkt=zh-cn) [B](http://cn.bing.com/dict/search?q=B&FORM=BDVSP6&mkt=zh-cn).[C](http://cn.bing.com/dict/search?q=C&FORM=BDVSP6&mkt=zh-cn).), [known](http://cn.bing.com/dict/search?q=known&FORM=BDVSP6&mkt=zh-cn) as the ["](http://cn.bing.com/dict/search?q=%22&FORM=BDVSP6&mkt=zh-cn)[Oracle](http://cn.bing.com/dict/search?q=Oracle&FORM=BDVSP6&mkt=zh-cn)["](http://cn.bing.com/dict/search?q=%22&FORM=BDVSP6&mkt=zh-cn) and was unified in the Qin dynasty (221-207 B.C.). In modern times, the most important changes in 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.

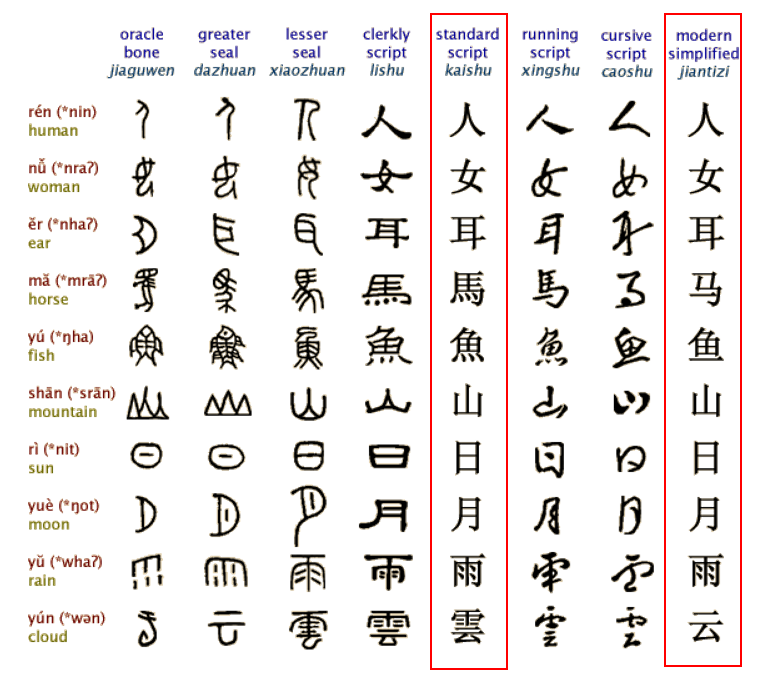


Figure 1: Evolution of Chinese Characters

As a result, the Chinese language has two sub 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. Hanzi in SC and TC share the same meaning and the same pronunciation and are typical variants.

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

The Chinese script spread 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.

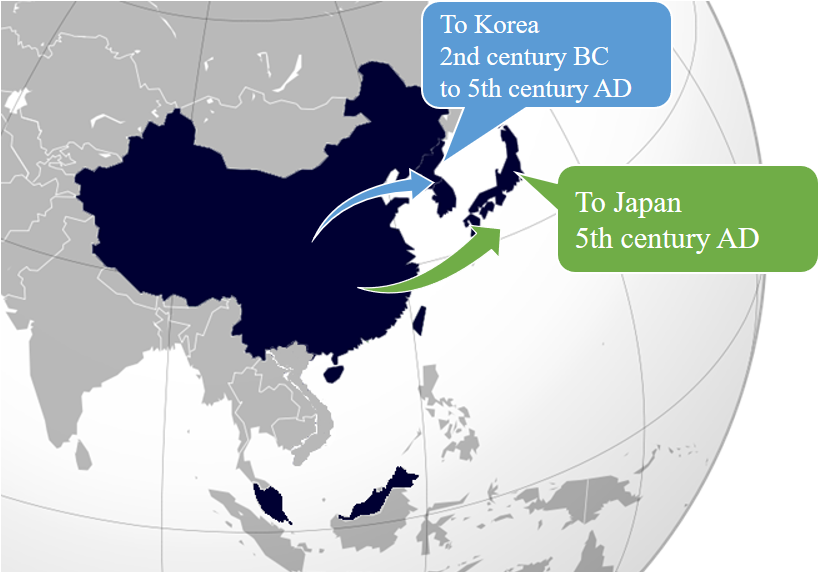
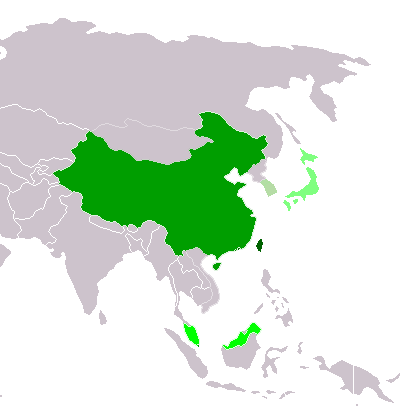


Figure 2: Chinese script spread to Japan and Korea

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 of the 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 languages 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 are used in Japanese in addition to two other scripts (hiragana and katakana), together known as Jpan (ISO 15924 code).
* Hanja are used in Korean in addition to the Hangul script, together known as Kore (ISO 15924 code).

The relationship among Hanzi (Hans & Hans), Kanji and Hanja is as shown below.

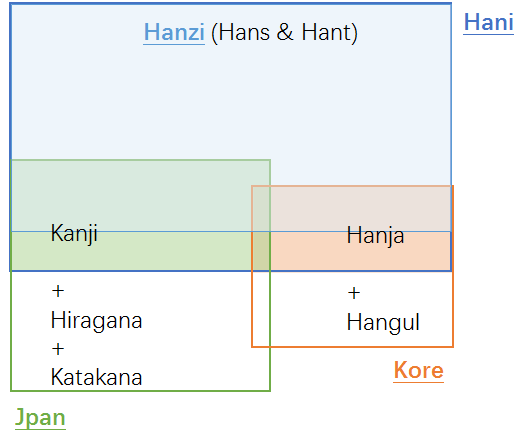


Figure 4: Hani, Hanzi, Kanji & Hanja

# Process for Developing the Proposal

## 2.1 Previous work

In April 2004, the Joint Engineering Team (JET), a group composed of members of CNNIC, TWNIC, KRNIC, 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, the 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. The **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, dotAsia, the registry of .ASIA and the member of CDNC, extended the CDNC IDN Table by importing characters from HKSCS (Hong Kong Supplementary Character Set) and the Singapore set, developed its own IDN table under the framework of CDNC rules, to cover needs from the Hong Kong and Singaporean local communities.

There has already been a detailed analysis of the 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.

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

Besides, the CGP is pleased to have **Edmon CHUNG**, CEO of dotAsia and Co-Chair of the 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 the JGP and KGP, to coordinate the Chinese code points and variant characters among three parties. The three Parties held three joint face-to-face meetings, in May 2015, March 2016 and November 2016, and had successive CJK joint sessions in ICANN meetings since ICANN 51 Los Angeles.

The work process includes the following steps:

* **Define and finalize the code point repertoire**

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

In addition to the CDNC IDN Table, there are some other character sets taken into account, including the dotAsia IDN Table, and the Table of General Standard Chinese Characters (TGSCC) published by China State Council.

The characters in the JGP repertoire and KGP repertoire are also referred to.

* **Define and finalize the code point variant sets**

CDNC has provided a market-proven variant set in the CDNC IDN table. Following CDNC rules, dotAsia extended CDNC repertoire and variant set to meet the requirements from the Hong Kong and Singaporean local communities. The CGP adopted CDNC variant rules directly and then made necessary updates related to dotAsia variant rules.

The 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. The CGP would work 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 the whole label evaluation ruleset**

The CGP WLE follows the spirit of the 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.

The CGP also acknowledges that some multiple preferred variant mappings work for SLD but may overproduce allocatable labels in the root zone. The CGP worked together with J, K and the IP to design an ideal solution to set applicants’ preferred labels allocatable as well as to limit the amount of allocatable labels to a reasonable number (for example, three).

* **Create XML LGR for Chinese script LGR proposal**

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

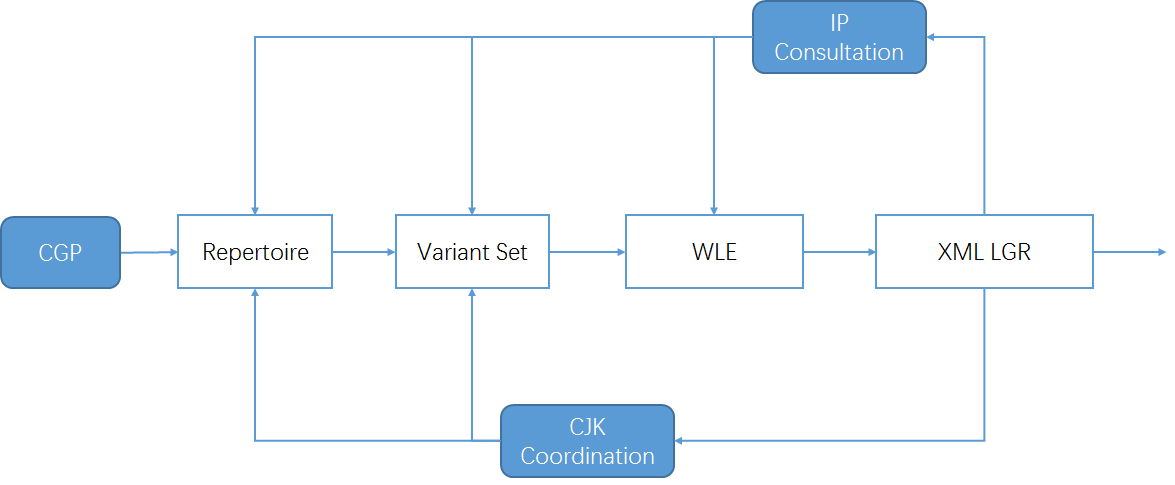


Figure 5: Iteration model of CGP work process

# Code point repertoire

## 3.1 Basic character set

**In 2004**, according to RFC 3743 and RFC 4713, the Chinese Domain Name Consortium (CDNC) drafted CDNC Chinese IDN Table. The CDNC Table has been used for second level domain (SLD) name registration under .CN, .TW, .HK and many CDN TLDs. In March, 2005, CNNIC and TWNIC submitted .CN Chinese Character Table[[1]](#footnote-1) and .TW Chinese Character Table[[2]](#footnote-2) separately, which included repertoire and variant mappings information.

**In 2012**, CDNC reviewed, proofread and published its combined IDN Table for the implementation of Chinese IDN registrations at gTLDs, including 37 ASCII code points and 19520 Chinese characters (<http://www.cdnc.org/gb/research/file/CDNC_unicode.txt>).

## 3.2 Repertoire formation process

### 3.2.1 Basic Repertoire

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).

In October 2015, CDNC published the latest version of it IDN Table; 41 new Chinese characters were added into the character set as requested by HKIRC on behalf of the 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 these **19561** code points form the basic set of the CGP repertoire. (CGP R0=CDNC IDN Table)

### 3.2.2 124 dotAsia characters

dotAsia extended CDNC IDN Table 2012, adding 163 additional 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 dotAasia 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 remaining 124 extended the CGP repertoire up to **19685** code points (CGP R1, covering the whole dotAsia IDN Table).

Among these 124 code points, only 62 are located in the Basic Multilingual Plane, the other 62 code points are from Supplementary Ideographic Plane (Plane 2) as below:

|  |  |  |  |
| --- | --- | --- | --- |
| Unicode | Character | Unicode | Character |
| 2070E | 𠜎 | 210C9 | 𡃉 |
| 20731 | 𠜱 | 211D9 | 𡇙 |
| 20779 | 𠝹 | 220C7 | 𢃇 |
| 20C53 | 𠱓 | 227B5 | 𢞵 |
| 20C78 | 𠱸 | 22AD5 | 𢫕 |
| 20C96 | 𠲖 | 22B43 | 𢭃 |
| 20CCF | 𠳏 | 22BCA | 𢯊 |
| 20CD5 | 𠳕 | 22C51 | 𢱑 |
| 20D15 | 𠴕 | 22C55 | 𢱕 |
| 20D7C | 𠵼 | 22CC2 | 𢳂 |
| 20D7F | 𠵿 | 22D08 | 𢴈 |
| 20E0E | 𠸎 | 22D4C | 𢵌 |
| 20E0F | 𠸏 | 22D67 | 𢵧 |
| 20E77 | 𠹷 | 22EB3 | 𢺳 |
| 20E9D | 𠺝 | 23CB7 | 𣲷 |
| 20EA2 | 𠺢 | 244D3 | 𤓓 |
| 20ED7 | 𠻗 | 24DB8 | 𤶸 |
| 20EF9 | 𠻹 | 24DEA | 𤷪 |
| 20EFA | 𠻺 | 2512B | 𥄫 |
| 20F2D | 𠼭 | 26258 | 𦉘 |
| 20F2E | 𠼮 | 267CC | 𦟌 |
| 20F4C | 𠽌 | 269F2 | 𦧲 |
| 20FB4 | 𠾴 | 269FA | 𦧺 |
| 20FBC | 𠾼 | 27A3E | 𧨾 |
| 20FEA | 𠿪 | 2815D | 𨅝 |
| 2105C | 𡁜 | 28207 | 𨈇 |
| 2106F | 𡁯 | 282E2 | 𨋢 |
| 21075 | 𡁵 | 28CCA | 𨳊 |
| 21076 | 𡁶 | 28CCD | 𨳍 |
| 2107B | 𡁻 | 28CD2 | 𨳒 |
| 210C1 | 𡃁 | 29D98 | 𩶘 |

### 3.2.3 18 TGSCC characters

After CDNC revealed Chinese Character Table version 2012, in 2013, China's State Council published **Table of General Standard Chinese Characters通用规范汉字表** (TGSCC)[[3]](#footnote-3) as  
<http://www.gov.cn/zwgk/2013-08/19/content_2469793.htm> (8105 characters).  
To stay consistent, the CGP studied the TGSCC and found 18 characters outside CGP R1 that fall in the range of the MSR. The CGP imported these 18 characters into the repertoire, making the total number 19703. (CGP R2)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | TGSCC | 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 |  |  |  |  |  |  |  |  |

### 3.2.4 43 CJK Coordination characters

Since the meaning is inherent to the symbol, the same logographic system can theoretically be used to represent different languages like Chinese, Japanese and Korean.

In the early 2000s, CDNC experts developed CDNC IDN Table (CGP R0) in which some IICORE Hani characters used in the Japanese or Korean writing systems are excluded. Under the framework of CJK coordination, the Hani characters included in the JGP repertoire or KGP repertoire need to be reviewed as regards the variant relationships to the CGP repertoire, to reach a tripartite consensus on the characters and variant mappings.

To ensure that the CGP repertoire will not bring any confusion or conflict to global Chinese character users and applicants at the root level, the CGP reviewed 99 IICORE characters in the MSR but not covered in CGP R2, and found 43 characters included in the JGP repertoire (version 201703, Appendix C) and KGP repertoire (version 201703, Appendix D) with variant relationships 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 |

The CGP imported these 43 characters into the repertoire, raising the total number to **19746.** (CGP R3)

The CGP repertoire and its component character sets are illustrated in the following figure:

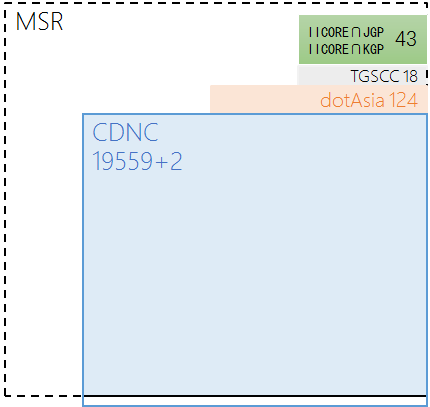


Figure 6: CGP R3 repertoire components

The CGP R3 repertoire totals 19735 Hanzi charaters.

The JGP repertoire (version 201703, Appendix C) includes 6358 Hani/Kanji characters. 6267 are overlapped characters in the CGP R3 repertoire. CGP will treat the other 91 as Japanese UNIQUE Kanji characters and will not seek to import any of them into the CGP repertoire. The JGP agreed to follow the CGP variant mappings of the overlapping 6267 characters as illustrated in Section 4.2.5.

The KGP repertoire (version 201703, Appendix D) has 4758 Hanja characters. 4749 are overlapping characters in the CGP R3 repertoire. The JGP agreed to follow the CGP variant mappings of the overlapping 4749 characters as illustrated in Section 4.2.5.

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

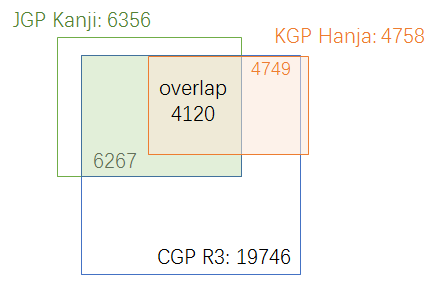


Figure 7：CGP Hanzi set, JGP Kanji set and KGP Hanja set

## 3.3 Attempt to limit the size of the repertoire

In Section 3.2, the CGP generated a repertoire containing 19,746 code points / characters. It is remarkable that the CGP repertoire has such a large size compared with most other GPs. CGP would attribute it to the nature of the Chinese writing system, similar to other logographic writing systems with large repertoires.

Unlike a segmental writing system (e.g. alphabetic, Abjad, Abugida) which has limited [graphemes](https://en.wikipedia.org/wiki/Grapheme) to represent the [phonemes](https://en.wikipedia.org/wiki/Phoneme) (basic units of sound) of a language, or Syllabaries (e.g. Kana) which has limited graphemes to represent [syllables](https://en.wikipedia.org/wiki/Syllable) or [moras](https://en.wikipedia.org/wiki/Mora_(linguistics)" \o "Mora (linguistics)), a logographic writing system has glyphs/logograms to represent [words](https://en.wikipedia.org/wiki/Word_(linguistics)) or [morphemes](https://en.wikipedia.org/wiki/Morpheme) rather than phonetic elements. In Chinese, a logogram is a single written character which represents a complete grammatical word (or, more precisely, a [morpheme](https://en.wikipedia.org/wiki/Morpheme)). As each character represents a single word, many logograms are required to write all the words of the language.

There are two reasons to explain why there are so many characters in the Chinese writing system. First, each Chinese character is an independent unit representing a word. 3000 years ago, the Shang [Dynasty](http://cn.bing.com/dict/search?q=Dynasty&FORM=BDVSP6&mkt=zh-cn) (16[th](http://cn.bing.com/dict/search?q=th&FORM=BDVSP6&mkt=zh-cn)-11[th](http://cn.bing.com/dict/search?q=th&FORM=BDVSP6&mkt=zh-cn) [century](http://cn.bing.com/dict/search?q=century&FORM=BDVSP6&mkt=zh-cn) [B](http://cn.bing.com/dict/search?q=B&FORM=BDVSP6&mkt=zh-cn).[C](http://cn.bing.com/dict/search?q=C&FORM=BDVSP6&mkt=zh-cn).) oracle bones included 3500 ~ 4500 characters already. During the course of history, more characters were invented to represent new words created along with social development. Second, massive numbers of variants occurred with the spread of Chinese characters and and the development of written communication in the continent of east Asia. **Chinese variants are characters with different visual forms but with the same pronunciations and with the same meanings as the corresponding official forms.** In the Chinese writing system, variants are deemed as exchangeable, the classic case is simplified characters and traditional characters. Generally, each Chinese character has at least one non-reflexive variant character (for CDNC IDN Table, average 1 non-reflexive variant, at most 7 non-reflexive variants).

Statistically, “[Text](http://cn.bing.com/dict/search?q=Text&FORM=BDVSP6&mkt=zh-cn) [Notes](http://cn.bing.com/dict/search?q=Notes&FORM=BDVSP6&mkt=zh-cn) [and](http://cn.bing.com/dict/search?q=and&FORM=BDVSP6&mkt=zh-cn) [Word](http://cn.bing.com/dict/search?q=Word&FORM=BDVSP6&mkt=zh-cn) [Explanations](http://cn.bing.com/dict/search?q=Explanations&FORM=BDVSP6&mkt=zh-cn) 说文解字/說文解字” in Han Dynasty (202 B.C.-220 A.D.) includes 9353 characters, “Lei Pian 类篇/類篇” in Song Dynasty (960-1279 A.D.) includes 31319 characters. In 1710, [Emperor](http://cn.bing.com/dict/search?q=Emperor&FORM=BDVSP6&mkt=zh-cn) [Kangxi](http://cn.bing.com/dict/search?q=Kangxi&FORM=BDVSP6&mkt=zh-cn) released the “Kangxi [Dictionary](http://cn.bing.com/dict/search?q=Dictionary&FORM=BDVSP6&mkt=zh-cn) 康熙字典” including 47035 characters. In 1959, the Japanese scholar Tetsuji Morohashi compiled “Dai Kan-Wa Jiten大汉和辞典” covering 49964 characters. In 1994, the Chinese Zhonghua Book Company published “Zhonghua Zihai 中华字海” containing 87019 characters. In 2004, the [Taiwan Ministry of Eduction released](http://dict.variants.moe.edu.tw/main.htm)“[Dictionary of Chinese Character Variants異體字字典”](http://dict.variants.moe.edu.tw/main.htm)containg 106,230 characters.

It is obvious that, among the tens of thousands of Chinese characters, not all are frequently used in modern society. The Chinese Ministry of Education requires that the students must handle 3500 characters after [nine](http://cn.bing.com/dict/search?q=nine&FORM=BDVSP6&mkt=zh-cn) [year](http://cn.bing.com/dict/search?q=year&FORM=BDVSP6&mkt=zh-cn)s of [compulsory](http://cn.bing.com/dict/search?q=compulsory&FORM=BDVSP6&mkt=zh-cn) [education](http://cn.bing.com/dict/search?q=education&FORM=BDVSP6&mkt=zh-cn), the number is 3500-4500 in Taiwan and 3500 in Hong Kong. However, everyday Chinese script users are able to “write” and “read” many more characters than what they actually learned in school due to two reasons.

The first reason is that Chinese variant characters have same pronunciation. Because of that, modern internet users who have received compulsory education prefer to use phonetic-based input methods (e.g. Pinyin拼音 in China mainland, Zhuyin注音 in Taiwan, Jyutping粤拼 in Hong Kong), which allow users to input [phonetic](http://cn.bing.com/dict/search?q=phonetic&FORM=BDVSP6&mkt=zh-cn) [symbol](http://cn.bing.com/dict/search?q=symbol&FORM=BDVSP6&mkt=zh-cn)s and select characters/labels from the alternative variant characters/labels with the same pronunciation in the selection box. Moreover, a few users prefer other input methods like shape-based input methods (e.g. Wubi五笔 in China, Simplified Tsang-jei 速成 in Hong Kong), handwriting recognition or speech recognition, however, most of them provide a phonetic-based selection box as a basic function to enable users to input variants with [no](http://cn.bing.com/dict/search?q=no&FORM=BDVSP6&mkt=zh-cn) [barriers](http://cn.bing.com/dict/search?q=barriers&FORM=BDVSP6&mkt=zh-cn).

The second reason is, generally, a set of Chinese variant characters share the same radical or components, and thus have a certain degree of visual similarity, allowing educated readers to recognize the variant relationship easily. For example, the character for “fight” has 6 variants with similar visual forms, 鬪(9B2A)闘(95D8)鬥(9B25)鬦(9B26)鬬(9B2C)鬭(9B2D). More importantly, almost every variant character does not appear alone in any domain label, but together with other characters in a word or phrase, providing semantic context and helping the readers to recognize the meaning of domain labels more effectively and conveniently. (For example 头发/头髪 ‘hair’ and 发展/發展 ‘development’)

The above two natural characteristics give Chinese variant characters great acceptability, usability and exchangeability in real life, especially in information systems. Hence, the [developement](http://cn.bing.com/dict/search?q=developement&FORM=BDVSP6&mkt=zh-cn) and popularization of the [internet](http://cn.bing.com/dict/search?q=internet&FORM=BDVSP6&mkt=zh-cn) promote Chinese character usage in cyberspace. In terms of [Computer](http://cn.bing.com/dict/search?q=Computer&FORM=BDVSP6&mkt=zh-cn) [Coding](http://cn.bing.com/dict/search?q=Coding&FORM=BDVSP6&mkt=zh-cn) [Standards](http://cn.bing.com/dict/search?q=Standards&FORM=BDVSP6&mkt=zh-cn), the early Taiwan BIG5 standard includes 13053 characters, the current Taiwan state standard CNS11643(4.0) includes 76,067 characters. China GB2312 standard included 6763 characters, while the latest standard GB18030 included 20912 characters. The current Unicode standard (as of 10 October 2015), including CJK Unified Ideographs Extensions A-E, contains 80,388 characters. In terms of internet application and daily usage, in 2007, the paper “A Survey on the Usage of Chinese Characters and Phrases in the Newspapers, Radio, TV, and Web” in Applied Linguistics [1003-5397(2007)01-0029-09] shows 8128 independent characters are used in daily life. Another paper in 2010, “survey on Chinese Weblog Wording” in Journey of Xianning University [1006-5342(2010)01-0076-03], shows 20923 characters are used.

The most symbolic event is, in 2016, China’s [Ministry](http://cn.bing.com/dict/search?q=Ministry&FORM=BDVSP6&mkt=zh-cn) [of](http://cn.bing.com/dict/search?q=of&FORM=BDVSP6&mkt=zh-cn) [Civil](http://cn.bing.com/dict/search?q=Civil&FORM=BDVSP6&mkt=zh-cn) [Affairs](http://cn.bing.com/dict/search?q=Affairs&FORM=BDVSP6&mkt=zh-cn) issued Notification 2016[33], requiring government departments to update the naming related information system in public service and administration areas, to cover the characters in national standard GB13000 (20902 chars) or GB18030 (70244 chars). The two standards cover the CGP repertoire entirely.  
http://www.gov.cn/xinwen/2016-05/09/content\_5071481.htm

Actually, all the above concerns were taken into account when the CDNC generated its Chinese IDN Table in the early 2000s. To create an IDN Table with broad applicability and [backwards](http://cn.bing.com/dict/search?q=backwards&FORM=BDVSP6&mkt=zh-cn) [compatibility](http://cn.bing.com/dict/search?q=compatibility&FORM=BDVSP6&mkt=zh-cn), the CDNC referred to multiple source files about Chinese characters and variants, including:

1. Complete List of Simplified Characters简化字总表 (2235 chars)
2. List of Commonly Used Characters in Modern Chinese现代汉语通用字表 (7000 chars)
3. China National Standard GB2312 (6763 chars)
4. Taiwan standard BIG5 (13053 chars)
5. Chinese Variants Collation Table 第一批异体字整理表 (810 variant groups)
6. Chinese Big Dictionary 汉语大字典 (54678 chars)
7. Chinese Relationship Table for Unihan Project
8. International Standard Chinese Big Dictionary国际标准汉字大辞典
9. Unicode 3.2
10. Unihan Database and extension A (20992 + 6582 chars)

The CDNC took Reference 1–- Reference 4 as sources to set up a fundamental character set, then imported variant characters from Reference 5 – Reference 8 to develop variant mappings, generated the CDNC IDN table with 19520 Chinese characters. All fall in the range of Reference 9 (Unicode 3.2) and Reference 10 (Unihan Database and extension A). In CDNC IDN Table, there are only 52 characters not included in GB13000, but all are included in GB18030.

In the early stage of developing the GP repertoire, CGP members tried to replace the CDNC IDN table with a smaller character set, hoping the reduction would 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) of the LGR and speed up the coordination work with J & K. The CGP generated a reduced repertoire called MSS (Minimum Shared Set) of 12563 characters, most of them are historically registered in SLD under .CN/.TW/.HK/.网址 (7722 chars) or come from the Table of General Standard Chinese Characters of the PRC State Council (4612 chars).

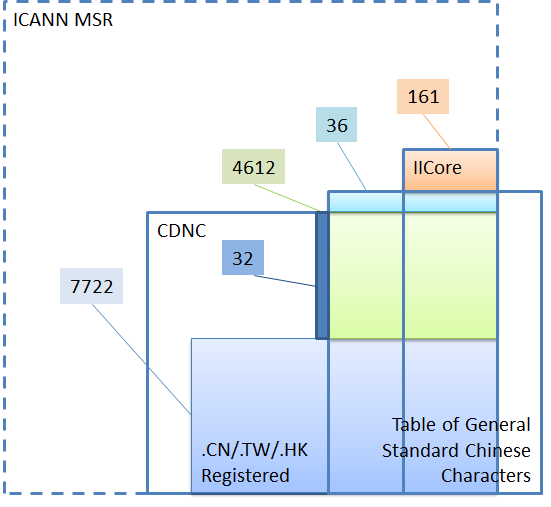


Figure 8: MSS constitution diagram

The CGP generated the MSS and expected that this limited repertoire could significantly decrease the complexity and workload of coordination between CJK, however, this reduction attempt caused a heated discussion among the CGP members, especially for those registry representatives who had already adopted the CDNC IDN Table for second level registrations. The core issue is that many members tend to believe that it is the variant mapping rules, not the repertoire size that directly affects the computational complexity of the LGR. The storage capacity and processing power of the modern computer is much more than what is needed to deal with a repertoire of about 20000 characters. Since the 2000s, many IDN registries adopted the CDNC IDN Table and developed IDN registration systems without decreasing the computational performance of the EPP service. Considering the SLD market acceptance of the existing CDNC IDN Table (adopted by over 5 ccTLDs and 20 new gTLDs) and the continuity of registries’/registrars’/registrants’ experience, many CGP members suggested the inheritance of the characters of the CDNC IDN Table to the maximum extent.

Moreover, CJK coordination work shows that the JGP has no dissent with the CGP repertoire and variant mappings. The KGP has no dissent with the CGP repertoire either, but only concerns about the mapping relationships of specific 258 variant groups.

For [all](http://cn.bing.com/dict/search?q=all&FORM=BDVSP6&mkt=zh-cn) the above [reasons](http://cn.bing.com/dict/search?q=reasons&FORM=BDVSP6&mkt=zh-cn)[,](http://cn.bing.com/dict/search?q=%2C&FORM=BDVSP6&mkt=zh-cn) the CGP decided to keep CGP R3 as the final repertoire with high capacity and [compatibility](http://cn.bing.com/dict/search?q=compatibility&FORM=BDVSP6&mkt=zh-cn), which means, to implement the consistency of user experiences, registry practices as well as the current local regulations .

# Code point variants

## 4.1 Variant definition in CGP

In the 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 form, but treated equally with universal interchangeability. This relationship of interchangeability is much stronger than the relationship between the Traditional and Simplified forms.

In the 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 inherited in the current CGP LGR document.

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

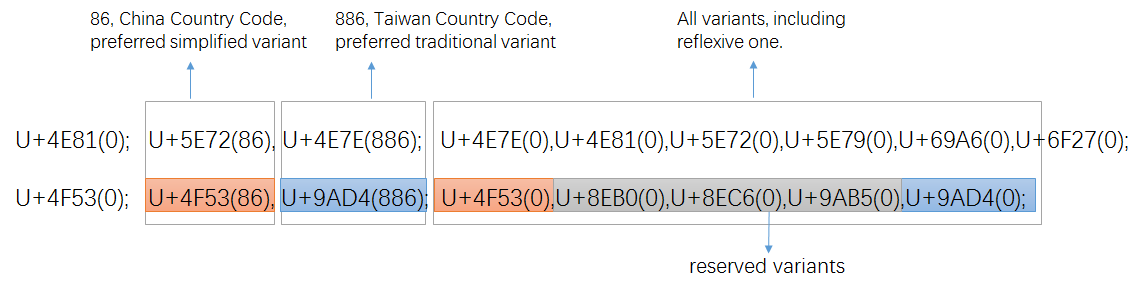


Figure 9: variant setting in CDNC IDN Table

Once transformed into XML-format (draft-davies-idntables-10, Representing Label Generation 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 Variant Mappings formation process

### 4.2.1 Basic Variant mappings

Considering that the majority of the CGP repertoire is imported from the CDNC IDN Table directly, the CGP borrowed the variant mappings in the CDNC IDN Table directly and developed the basic variant mappings table corresponding to CGP R0.

### 4.2.2 172 TGSCC and IICORE variants review

In the 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 are included in CGP R3, and could be variants of previous CGP characters. To ensure that the CGP repertoire will not bring any confusion or conflict to global Chinese character users and applicants at the root level, the CGP & CDNC held joint meetings and invited linguistic experts from China mainland, Taiwan and Hong Kong to review 172 Hanzi characters (15 TGSCC characters and 157 IICORE Hanzi) outside CGP R0.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **Hong Kong** | **TGSCC** | **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 CGP and CDNC held a joint meeting and invited experts from China mainland, Taiwan and Hong Kong to review all 172 characters and output a variant mapping review document as **Appendix G**. Based on this variant mapping review document, the CGP imported necessary variant characters and adjusted the related variant mappings.

### 4.2.3 69 dotAsia unique variants review

In Haikou, in May 2016, the CGP & CDNC joint meeting reviewed 7 dotAsia unique Hanzi characters. These Hanzi characters are not included in the CDNC IDN table, nor in TGSCC, nor in IICORE they they only exist in the dotAsia IDN table submitted to IANA.   
http://www.iana.org/domains/idn-tables/tables/asia\_zh\_1.1.txt



The variant mappings of the 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):

* Merge correlative variant characters into a union set
* Reset the preferred-simp and preferred-trad for the 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) | 𨋢(282E2) |  | 䢂(4882) | 𨋢(282E2) |  |
| 4C9D | 䲝 | dotAsia | 䲝(4C9D) | 䱽(4C7D) |  | 䲝(4C9D) | 䱽(4C7D) | 鲳(9CB3) 鯧(9BE7) |
| 4C9E | 䲞 | dotAsia | 䲞(4C9E) | 𩶘(29D98) |  | 䲞(4C9E) | 𩶘(29D98) |  |
| 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) |

Moreover, in the dotAsia IDN table, there are 62 IICORE code points from Unicode Plane 2. The CGP intends to directly accept their variant mappings from dotAsia IDN Table into CGP rules.

The CGP and Edmon CHUNG discussed the issue of inconsistency between the CGP variant mappings and dotAsia variant mappings, and agreed that the dotAsia table was created as an experiment for Hong Kong local characters, but the intent has always been to merge it and make it consistent with CGP rules once it is integrated for root zone and gTLD purpose. Thus, dotAsia agreed to synchronize and update the IDN table in IANA once the CGP rules are finalized.

### 4.2.4 CJK Variant coordination

#### 4.2.4.1 Principle and Framework

A coordination mechanism among three parties is needed to realize unified Chinese script generation rules in the DNS root zone. During the CDNC meeting in Shanghai (May, 2014), the 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, the CGP, JGP and KGP started coordination work at the IETF Dallas meeting in 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 at the IETF Dallas meeting, the JGP initiated 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 merges WLE in LGR-alpha into one.

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

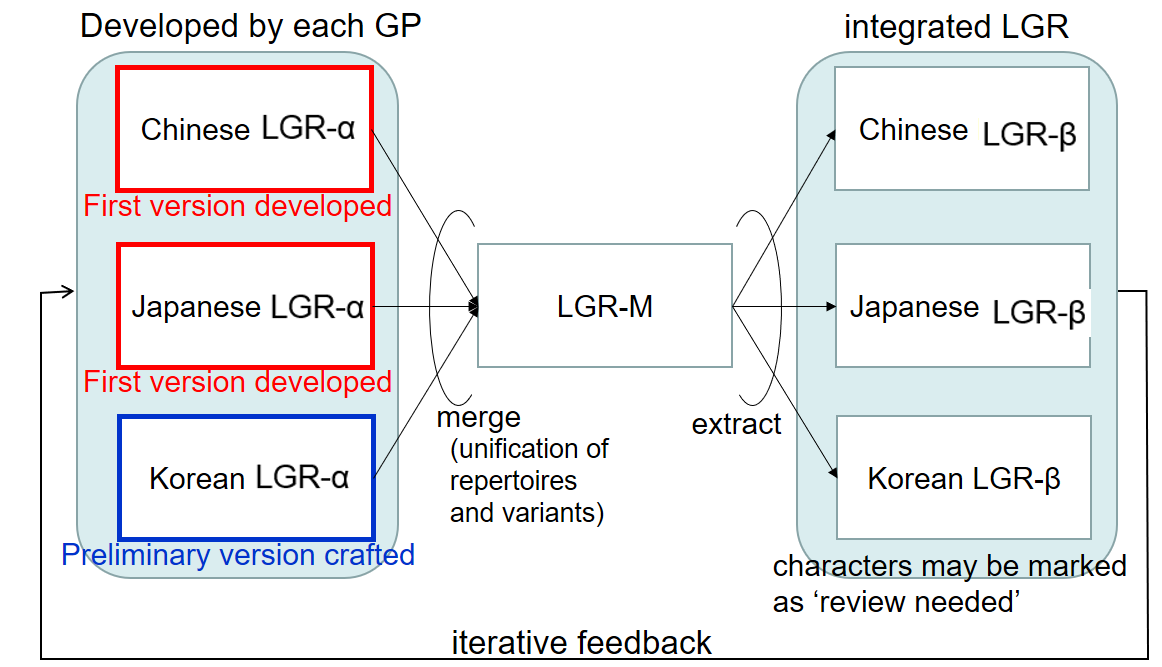


Figure 10: Framework of CJK LGR integration for Han characters, by JGP

#### 4.2.4.2 Coordination between C and J

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 the 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.

Some characters in a CGP variant group have the same pronunciations and meanings, but have different meanings in Japanese language environments. For example, (U+673机) means [desk, small table] and (U6A5F機) means [machine] in Japanese, but both mean [machine] in Chinese.

The JGP showed great openness and agreed to import all CGP variant mappings into the JGP ones. Thus, both parties eliminated the potential conflict caused by variant inconsistency. The CGP would like to express its appreciation for the JGP’s openness, tolerance and compromise.

#### 4.2.4.3 Coordination between C and K

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 are still sometimes used by a few Korean people in daily life. In August 2016, the KGP generated its first version of its LGR, and raised 258 Hanja/Hanzi variant groups whose variant mappings in the CGP LGR are NOT acceptable to the KGP (Appendix F).

Unlike the JGP, the KGP adopted another strategy and approach to resolve differences on variant mappings, asking the CGP to review and reset the variant mapping related to the 258 variant groups case by case.

As requested, the CGP made statistics and analysis of all the disputed 258 variant groups, including the number of registered labels containing disputed characters under .CN/.TW/.HK/.网址, and the semantics of all these registered labels. The CGP separated 258 variant groups into 5 types and 2 categories, the variant groups the CGP would RESERVE, the ones the CGP would give up and split into INDEPENDENT characters.

|  |  |  |
| --- | --- | --- |
| Reserved 1 | All registered number of variants are not zero, the actual meaning of the variants are the same, suggest to follow CGP rules | 139 |
| Reserved 2 | One variant’s registration number is almost zero, but the actual meaning of the variants are the same, suggest to follow CGP rules |
| Independent 1 | All registered numbers of variants are zero, not commonly used in domain registration, suggest dealing with independently | 119 |
| Independent 2 | All registered numbers of variants are not zero, meaning of variants are different, suggest dealing with independently |
| Independent 3 | One variant’s registration numbers are almost zero, not commonly used in domain registration, suggest dealing with independent |

The KGP invited linguistic experts from the local community and separated the 258 groups into 2 categories, 149 groups the KGP would KEEP and 109 groups the KGP would DROP.

|  |  |  |
| --- | --- | --- |
| Keep | Some CGP variant mappings don’t make any sense in the Korean Hangul environment, instead, these characters are used independently with their own semantic meanings. Hence, KGP intended to keep them as independent characters. | 149 |
| Drop | Some CGP variant mappings are weakly related in the Korean Hangul environment; the KGP intends to drop their independency and accept the CGP’s variant mappings. | 109 |

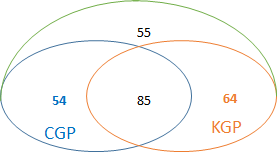


Figure 11: coordination before C&K Taiwan meeting

In September 2016, the KGP and CGP held the first round coordination meeting and reduced the number of variant groups in disagreement to 60. In the IETF meeting, in November 2017, the KGP and CGP reduced the number to 2. Finally, in December 2017, the KGP and CGP solved all 258 variant groups and reached complete agreement on the variant mappings. For the 258 unacceptable variant groups, 112 will go with CGP rules, the remaining 146 will go with KGP rules.

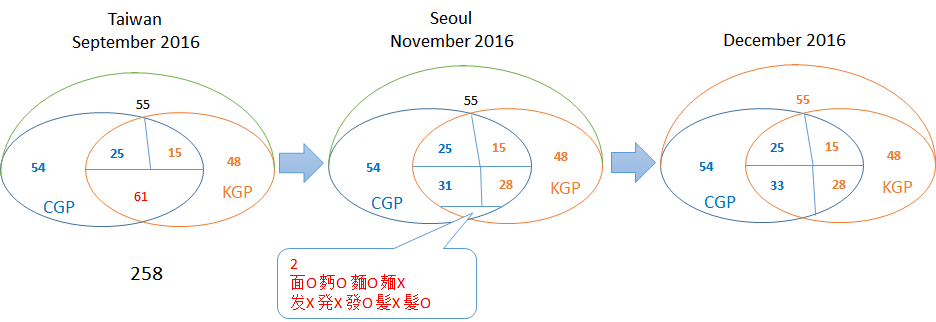


Figure 12: C&K coordination process

Consequently, the CGP updated 445 variant mappings related to 146 groups as **Appendix H.**

After the above steps, the whole CGP variant mapping table corresponding to CGP R3 was generated as **Appendix I**.

## 4.3 Multiple Preferred Variant Characters

In the Chinese writing system, many characters have at least one non-reflexive variant characters, or multiple variant characters. According to RFC 4743, if the multiple variant characters are reserved, they will not be included in the allocatable labels; but if the multiple variant characters happen to be preferred, then multiple allocatable labels will be output as allocatalbe labels. In RFC 4713, this scenario is described in more details, to address the issue caused by multiple preferred simplified variants (PSV) or traditional variants (PTV).

|  |  |  |  |
| --- | --- | --- | --- |
| Unicode | Original Char | Preferred Simplified Char | Preferred Traditional Char |
| 5925 | 夥 | 夥(5925) 伙(4F19) | 夥(5925) |
| 647A | 摺 | 摺(647A) 折(6298) | 摺(647A) |
| 9EBD | 麽 | 麽(9EBD) 么(4E48) | 麼(9EBC) |
| 5347 | 升 | 升(5347) | 升(5347) 昇(6607) 陞(965E) |
| 53F0 | 台 | 台(53F0) | 台(53F0 )檯(6AAF) 臺(81FA) |
| 590D | 复 | 复(590D) | 复(590D) 復(5FA9) 複(8907) |
| 5C40 | 局 | 局(5C40) | 局(5C40) 侷(4FB7) 跼(8DFC) |
| 5E76 | 并 | 并(5E76) | 并(5E76) 並(4E26) 併(4F75) |

Multiple preferred variant mapping examples

To avoid over production of multiple allocatable labels, the generation rules and system are designed with a limitation 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 the CGP variant mapping table (Appendix I), for all 19746 characters, there are 3 characters with 2 PSVs, 127 with 2 PTVs, 5 with 3 PTVs and 1 with 4 PTVs, which means, all together 136 characters have multiple preferred variant characters. These 136 characters will generate multiple all-simplified labels or all-traditional labels, which will violate CDNC constraints and cause an over-production issue at the root zone level. In SLD practice, CDNC members designed a ranking selection function or human interaction mechanism to avoid the dilemma, to enable the applicants to SELECT at most one all-simplified and at most one all-traditional label 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 are reserved. Unlike the blocked labels defined in the LGR, these reserved allocatable labels might be activated at the request of an applicant later.

On the other hand, at the Root Zone level, without a “human interaction” mechanism, multiple variant character mappings would lead to overproduction of variant labels with an "allocatable" status.

An example would be “Taiwan” 台(53F0)湾(6E7E).

|  |  |  |  |
| --- | --- | --- | --- |
| Unicode | Original Char | Preferred Simplified Char | Preferred Traditional Char |
| 53F0 | 台 | 台(53F0) | 台(53F0) 檯(6AAF) 臺(81FA) |
| 6E7E | 湾 | 湾(6E7E) | 灣(7063) |

An input of 台(53F0)湾(6E7E) results in 3 allocatable traditional labels:

Original: 台(53F0) 湾(6E7E)

All-Simplified: 台(53F0) 湾(6E7E)

All-Traditional: 台(53F0) 灣(7063), 檯(6AAF) 灣(7063), 臺(81FA) 灣(7063)

Actually, except for台(53F0)灣(7063), the other two traditional combinations are either meaningless from a semantic standpoint or rarely used in daily life.

According to the IP’s suggestion in the 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 preferred simp/trad mappings**. The problem is that it is hardly to determine which one should be retained and which one should be removed. Many scenarios need [specific](http://cn.bing.com/dict/search?q=Specific&FORM=BDVSP6&mkt=zh-cn) semantic [combinations](http://cn.bing.com/dict/search?q=combinations&FORM=BDVSP6&mkt=zh-cn), rather than random preferred simp/trad labels.

An example would be “counter/bar” 柜(67DC) 台(53F0).

|  |  |  |  |
| --- | --- | --- | --- |
| Unicode | Original Char | Preferred Simplified Char | Preferred Traditional Char |
| 67DC | 柜 | 柜(67DC) | 柜(67DC) 櫃(6AC3) |
| 53F0 | 台 | 台(53F0) | 台(53F0) 檯(6AAF) 臺(81FA) |

An input of 柜(67DC)台(53F0) results in 6 allocatable traditional labels:

Original: 柜(67DC) 台(53F0)

All-Simplified: 柜(67DC) 台(53F0)

All-Traditional: 柜(67DC) 台(53F0), 柜(67DC) 檯(6AAF), 柜(67DC) 臺(81FA)

櫃(6AC3) 台(53F0), 櫃(6AC3) 檯(6AAF), 櫃(6AC3) 臺(81FA)

Unlike the last case of 台(53F0) 湾(6E7E), whose desired traditional label uses 台(53F0), the desired traditional label of 柜(67DC) 台(53F0) uses 檯(6AAF), a different variant form with the component “木wood”, meaning platform made in wood.

The above two cases show how risky it is to eliminate multiple preferred simp/trad mappings. Without the semantic information of the whole label, reducing the multiple preferred variant characters arbitrarily could lead to the absence of the desired label.

The CGP fully understands the IP’s concern that multiple preferred variant mappings might result in another serious problem, the overproduction of allocatable labels. For example, if the applied label is long enough and contains N characters with 3 PTV, t[heoretically](http://cn.bing.com/dict/search?q=theoretically&FORM=BDVSP6&mkt=zh-cn), the generation rules might output allocatable traditional labels. It could be an unbearable number for the root zone management. However, the length of the applied for label and the number N are never really unlimited in the real world. According to the statistical data under .CN/.TW/.HK/.ASIA/.网址, when the CGP repertoire and variant mappings are applied, the numbers related to REAL registered original labels are as below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Max Length of Applied Label | Max Number of PSV or PTV | Max Number of Allocatable Simp Label | Max Number of Allocatable Trad Label |
| CN | 38 | 6 | 2 | 96 |
| TW | 26 | 5 | 2 | 243 |
| HK | 15 | 3 | 2 | 9 |
| MO | 6 | 2 | 1 | 4 |
| AISA | 17 | 2 | 1 | 4 |
| 网址 | 21 | 4 | 1 | 32 |

The statistical data show that TLDs have limited allocatable labels when applied with CGP variant mappings, CN and TW have a larger “max number of allocatable labels” compared to other TLDS.

For .CN, there are only 5 original labels whose alloacatable labels number over 64:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Applied Label | Length | Chars with PSV or PTV | Chars | Number of Allocatable Simp Label | Max Number of Allocatable Trad Label |
| 河南省延津县汇丰钢构彩板厂 | 13 | 6 | 汇丰构彩板厂 | 1\*1\*1\*1\*1\*1=1 | 2\*2\*2\*2\*2\*2=64 |
| 沈阳市苏家屯区霖丰珍珠岩制品厂 | 15 | 6 | 沈苏家丰岩厂 | 1\*1\*1\*1\*1\*1=1 | 2\*2\*2\*2\*2\*2=64 |
| 石家庄蒙商旅游开发有限公司 | 13 | 5 | 家庄蒙游发 | 1\*1\*1\*1\*1\*1=1 | 2\*2\*2\*2\*2\*2=64 |
| 菏泽开发区汇金源彩钢复合板厂 | 14 | 6 | 发汇彩复板厂 | 1\*1\*1\*1\*1\*1=1 | 2\*2\*2\*2\*2\*2=96 |
| 德州经济开发区聚丰彩钢复合板加工厂 | 17 | 6 | 发丰彩复板厂 | 1\*1\*1\*1\*1\*1=1 | 2\*2\*2\*2\*2\*2=96 |

For .TW, there are 41 original labels whose alloactable labels number over 64:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Applied Label | Length | Chars with PSV or PTV | Chars | Number of Allocatable Simp Label | Max Number of Allocatable Trad Label |
| Anonymouse Label | 9 | 4 | 局台局局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 18 | 4 | 台局局局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 19 | 4 | 台局局局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 16 | 4 | 台局局局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 17 | 4 | 台局局局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 18 | 4 | 台局局局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 19 | 4 | 台局局局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 17 | 4 | 台局局局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 18 | 4 | 台局局局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 19 | 4 | 台局局局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 9 | 4 | 台台局局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 9 | 4 | 台台局局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 10 | 4 | 台局台局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 14 | 4 | 台局台局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 12 | 4 | 台局台局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 12 | 4 | 台局台局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 12 | 4 | 台局台局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 12 | 4 | 台局台局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 9 | 4 | 局台局局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 18 | 4 | 台局局局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 19 | 4 | 台局局局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 16 | 4 | 台局局局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 10 | 4 | 台局台局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 14 | 4 | 台局台局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 12 | 4 | 台局台局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 12 | 4 | 台局台局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 12 | 4 | 台局台局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 12 | 4 | 台局台局 | 1\*1\*1\*1=1 | 3\*3\*3\*3=81 |
| Anonymouse Label | 15 | 5 | 台局松局松 | 1\*1\*1\*1\*1=1 | 3\*3\*2\*3\*2=108 |
| Anonymouse Label | 15 | 5 | 台局松局松 | 1\*1\*1\*1\*1=1 | 3\*3\*2\*3\*2=108 |
| Anonymouse Label | 15 | 5 | 台局松局松 | 1\*1\*1\*1\*1=1 | 3\*3\*2\*3\*2=108 |
| Anonymouse Label | 16 | 5 | 台局松局局 | 1\*1\*1\*1\*1=1 | 3\*3\*2\*3\*3=162 |
| Anonymouse Label | 17 | 5 | 台局松局局 | 1\*1\*1\*1\*1=1 | 3\*3\*2\*3\*3=162 |
| Anonymouse Label | 16 | 5 | 台局松局局 | 1\*1\*1\*1\*1=1 | 3\*3\*2\*3\*3=162 |
| Anonymouse Label | 17 | 5 | 台局松局局 | 1\*1\*1\*1\*1=1 | 3\*3\*2\*3\*3=162 |
| Anonymouse Label | 16 | 5 | 台局松局局 | 1\*1\*1\*1\*1=1 | 3\*3\*2\*3\*3=162 |
| Anonymouse Label | 17 | 5 | 台局松局局 | 1\*1\*1\*1\*1=1 | 3\*3\*2\*3\*3=162 |
| Anonymouse Label | 11 | 5 | 台局台局局 | 1\*1\*1\*1\*1=1 | 3\*3\*3\*3\*3=243 |
| Anonymouse Label | 11 | 5 | 台局台局局 | 1\*1\*1\*1\*1=1 | 3\*3\*3\*3\*3=243 |
| Anonymouse Label | 14 | 5 | 台局台局局 | 1\*1\*1\*1\*1=1 | 3\*3\*3\*3\*3=243 |
| Anonymouse Label | 14 | 5 | 台局台局局 | 1\*1\*1\*1\*1=1 | 3\*3\*3\*3\*3=243 |

The CDNC and CGP acknowledge the fact that multiple variant mappings might lead to a big number of allocatable labels, however, the CDNC and CGP tend not to address the issue simply by limiting the PSV or PTV number due to respect for the natural characteristics of Chinese variants as well as the end users’ cognition of and experience with them.

The first one is Chinese variant characters have the same pronunciation. Because of that, modern internet users who have received compulsory education prefer to use phonetic-based input methods (e.g. Pinyin拼音 in China mainland, Zhuyin注音 in Taiwan, Jyutping粤拼 in Hong Kong), which allow users to input [phonetic](http://cn.bing.com/dict/search?q=phonetic&FORM=BDVSP6&mkt=zh-cn) [symbol](http://cn.bing.com/dict/search?q=symbol&FORM=BDVSP6&mkt=zh-cn)s and select characters/labels from the alternative variant characters/labels with the same pronunciation in the selection box. Moreover, a few users prefer other input methods like shape-based input methods (e.g. Wubi五笔 in China, Simplified Tsang-jei 速成 in Hong Kong), handwriting recognition or speech recognition, however, most of them provide a phonetic-based selection box as a basic function to enable users to input variants with [no](http://cn.bing.com/dict/search?q=no&FORM=BDVSP6&mkt=zh-cn) [barriers](http://cn.bing.com/dict/search?q=barriers&FORM=BDVSP6&mkt=zh-cn).

The second one is generally a set of Chinese variant characters share the same radical or component, and thus have a certain degree of visual similarity, allowing educated readers to recognize the variant relationship easily. Moreover, almost every variant character does not appear alone in any domain label, but together with other characters in a word or phrase, helping the readers to recognize the meaning more effectively and conveniently. (For example 头发/头髪 ‘hair’ and 发展/發展 ‘development’).

|  |  |  |  |
| --- | --- | --- | --- |
| Unicode | Original Char | Preferred Simplified Char | Preferred Traditional Char |
| 5347 | 升 | 升(5347) | 升(5347)昇(6607)陞(965E) |
| 53F0 | 台 | 台(53F0) | 台(53F0)檯(6AAF)臺(81FA) |
| 590D | 复 | 复(590D) | 复(590D)復(5FA9)複(8907) |
| 5C40 | 局 | 局(5C40) | 局(5C40)侷(4FB7)跼(8DFC) |
| 5E76 | 并 | 并(5E76) | 并(5E76)並(4E26)併(4F75) |
| 8499 | 蒙 | 蒙(8499) | 蒙(8499)懞(61DE)濛(6FDB)矇(77C7) |

Table: examples of variant characters with visual similarity

The above two natural characteristics give Chinese variant characters great acceptability, usability and exchangeability in real life. Chinese internet users are used to applying different variant forms in different scenarios (e.g. simplified in the China mainland market, traditional in the Taiwan market, archaic forms in ancient buildings or galleries). [In view](http://cn.bing.com/dict/search?q=In%20view&FORM=BDVSP6&mkt=zh-cn) of [this](http://cn.bing.com/dict/search?q=this&FORM=BDVSP6&mkt=zh-cn) [situation](http://cn.bing.com/dict/search?q=situation&FORM=BDVSP6&mkt=zh-cn), the CGP would follow the CDNC’s practice, to keep all multiple variant characters in the LGR.

We would emphasize again that, 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. As compensation, some supplementary measures were introduced to ensure the applicant gets the desired variant label, for example a post-delegation activation process (activate reserved labels as requested). Similarly, the CGP would like to propose an experimental solution to reduce the number of allocatable labels without eliminating multiple mappings in the following two steps:

**STEP 1, multiple variant sub-type**

Instead of simply eliminating multiple mappings, CGP would like to create 6 new sub-types, to identify the variant characters in multiple mappings.

|  |  |  |
| --- | --- | --- |
| 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 |

The variant chars with these sub-types are not supposed to be treated simply as BLOCKED, but as RESERVED preferred ones. On the other hand, in practice, the label contains them will be set to BLOCKED. That way, all characters will have only one simplified variant and one traditional variant, and the new LGR could ultimately keep the number of allocatable labels under THREE.

**STEP 2, multiple LGR execution**

The negative impact of the above 6 sub-types is obvious too. They could prevent applicants from getting specific desired all-simplified or all-traditional labels. To eliminate the negative impact, a **compensatory solution** could be designed.

The solution could be executing the LGR MULTIPLE times (at most THREE times), allowing the applicant to input multiple variant applied-for labels and merging the multiple outputs (at most 9 labels), to ensure that the applicant will exactly get the desired labels allocated. The CGP noticed that the JGP has a similar concern and requirement for MULTIPLE LGR execution. The process takes one or more original input labels and generates allocatable/blocked/invalid labels as output, where original input labels are marked as ‘allocatable’ in the output.

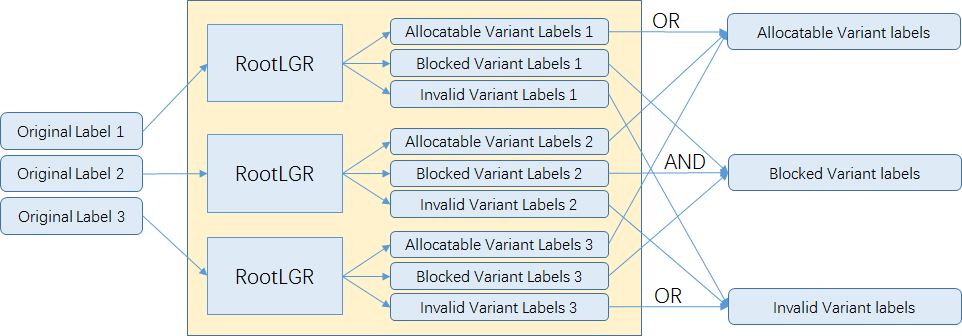


Figure 10: multiple LGR execution / complement process of root LGR from JGP

Alternatively, another solution could be setting up an extra process for applicants to request reserved label activation. The reserved labels are consisted of all chars with “r-simp-m”, “simp-m”, “r-trad-m”, “trad-m”, “r-both-m” or “both-m”.

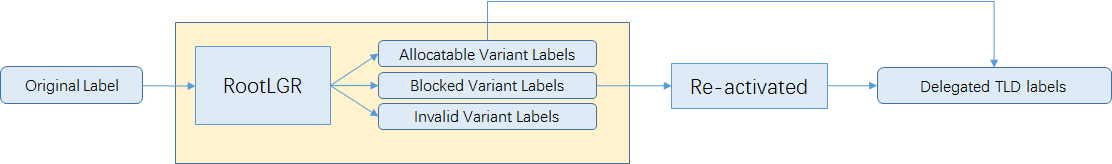


Figure 13: re-activation process

In CGP LGR 201703 (Appendix J), the CGP decided to introduce the 6 new subtypes (“simp-m”, “trad-m”, “r-simp-m”, “r-trad-m”, “r-both-m” and “both-m”) and, correspondingly, new rules for these sub-types into the LGR. But the new sub-types will not be actually tagged to any chars before either of the above two solutions was accepted by the IP.

# Whole Label Evaluation Rules

Following the XML-format transforming regulations (Representing Label Generation Rulesets using XML, <https://datatracker.ietf.org/doc/draft-davies-idntables/>), the 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, the 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” and “both-m”), but they don’t really work because no variant chars are tagged with these subtypes in the current Variant Mappings (Appendix J CGP Variant Mappings 201703).

# 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 Table of General Standard Chinese Characters (TGSCC)  
<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 Membership

# 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: 445 Variant Mappings Update

# Appendix I: CGP Variant Mappings

# Appendix J: CGP LGR XML

1. <http://www.iana.org/domains/idn-tables/tables/cn_zh-cn_4.0.html> [↑](#footnote-ref-1)
2. <http://www.iana.org/domains/idn-tables/tables/tw_zh-tw_4.0.1.html> [↑](#footnote-ref-2)
3. <https://en.wikipedia.org/wiki/Table_of_General_Standard_Chinese_Characters> [↑](#footnote-ref-3)