**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.

# Script and Languages Covered

## 1.1 Overview

Chinese characters are a sort of logograms 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.

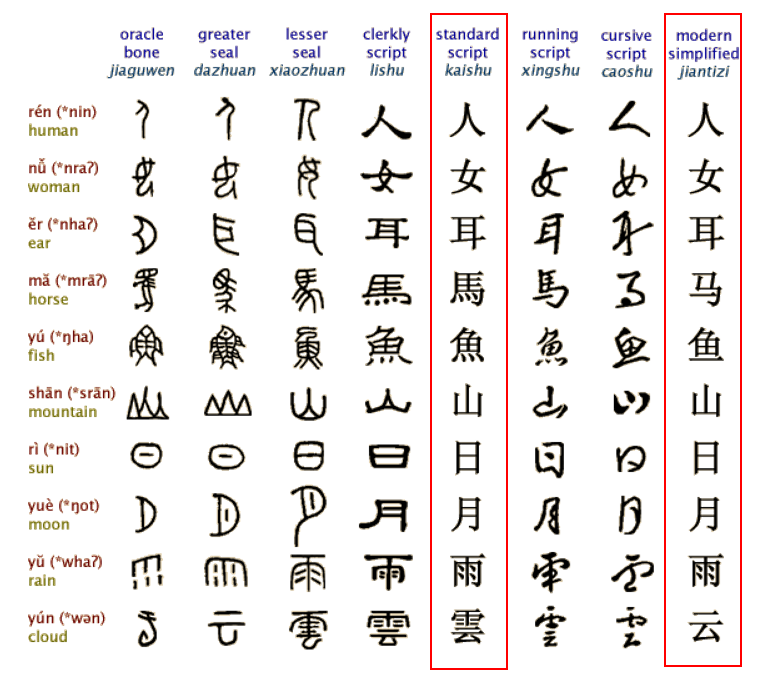


Figure 1: Evolution of Chinese Character

As a result, the Chinese language has got 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.

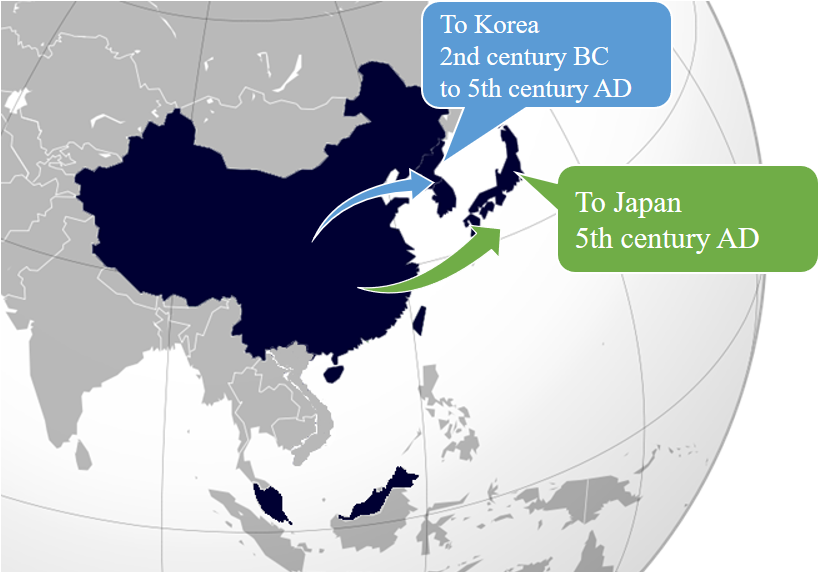


Figure 2: Chinese script spread to Japan and Korea

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:

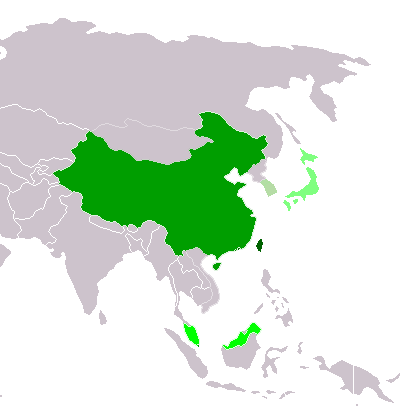


Figure 3: Countries using Chinese script

|  |  |
| --- | --- |
|  | 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) |

## 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 less significant.

For example, there are cases where a language may have a large population, but only a small part of it writes it 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 Hanjia |

* 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.

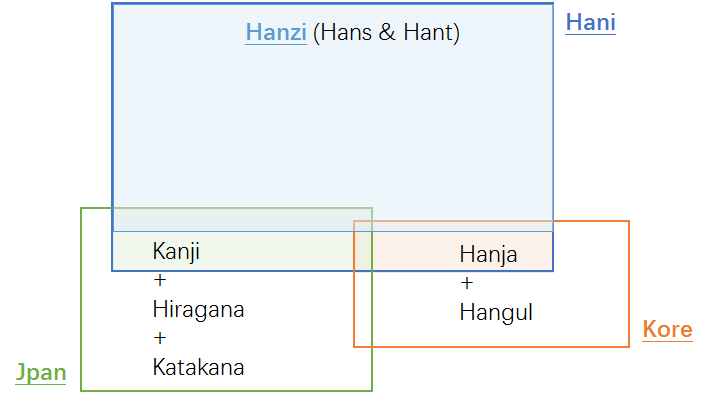


Figure 4: Hanzi, Kanji & Hanja

# 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 registrant. 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.

There has already been a detailed analysis of Chinse 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 Chinse 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 the experts from CDNC, who represents the Chinese language ccTLDs to a large extent, as well as the experts with a variety of backgrounds.

Geographically, the CGP has members from Chinese language regions across the 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.

## 2.3 Work Process

The work has been carried out since Sep 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 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 hold joint 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 are taken into account, including IICore (International Ideographs Core), Normalized Hanzi List for Common Use published by China State Council, and some New gTLDs applicant IDN tables.

* Define and finalize the code point variant sets

CDNC has given an industry-proven variant set in CDNC IDN table. But CGP recognizes that different panels (C, J and K) have different views on the variants corresponding to the same Chinese character, some CDNC’s variant mappings conflict with KGP and JGP’s perception and practice. CGP was working closely with JGP & KGP to make necessary compromise 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 registrant, while allocate the original applied label as well as only one preferred SC label and only one preferred TC label, totally no more than three labels, and blocks all other labels. Yet the XML-format transforming draft doesn’t illustrate how to generate a limited number of allocatable labels which contains registrant preferred labels. CGP is working together with JGP to devise an algorithm to reach that goal.

* Create XML LGR for Chinese script LGR proposal

Considering the fact that the coordination on variant mapping among CJK and the consultation on a limited number of allocatable labels with IP are still in the progress, the CGP work will be carried out in a fast iteration model as indicated in the following figure:

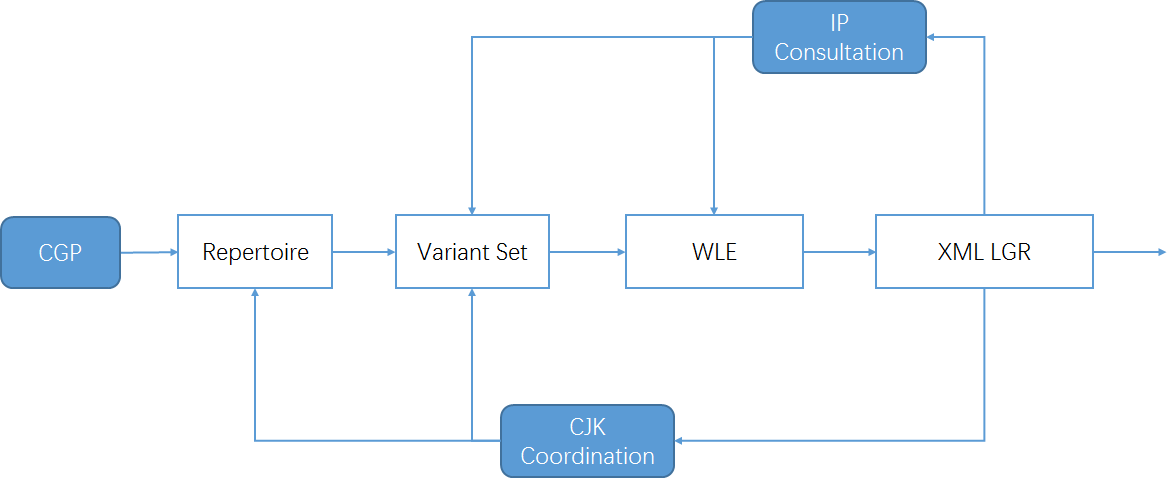


Figure 5: Iteration model of CGP work process

# 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 about 260, 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 registrants 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.

## 3.2 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). All those **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 community in 2013 and 2015, increasing the number of Chinese characters to 19561, but two of them are out of scope of MSR, which means only **19559** are included in MSR.
* In 2013, the China's State Council published Normalized Hanzi List for Common Use, CGP studied the Normalized List and found 27 characters fall in the range of MSR as well. CGP imported those 27 characters into the repertoire, making the total number **19586**.
* Furthermore, to ensure that CGP repertoire reserves enough character extensibility to meet the future registration demand from other Chinese script community in Asia-Pacific area, CGP studied IICore and imported 145 overlapped characters of IICore and MSR into repertoire, raising the total number up to **19731**.
* In December 2015, IP sent out the “Integration Pal Considerations on the November 25th, 2015 Chinese Root LGR”, in which seven more characters in dotAsia IDN table (<https://www.iana.org/domains/idn-tables/tables/asia_zh_1.1.txt>) were recommended to be covered, making CGP repertoire containing **19738** Chinese Characters eventually.

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

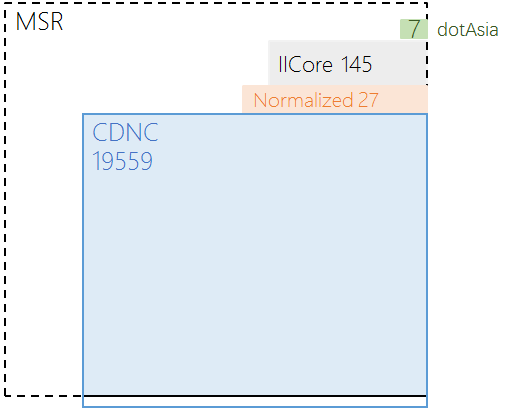


Figure 6: CGP repertoire components

## 3.2 Repertoire coordination with JGP and KGP

* For the Kanji characters in JGP repertoire, there are 6306 overlapped characters in CGP 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 the Hanja characters in KGP repertoire, all 4819 characters are included in CGP repertoire. If there are some Korean UNIQIE Hanja characters in the future, CGP will treat them in the same way as Japanese UNIQUE Kanji characters.

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

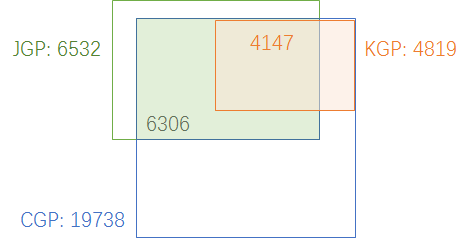


Figure 7：CGP Hanzi set, JGP Kanji set and KGP Hanja set

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.

For all 19738 characters in CGP repertoire, in addition to the original variant mappings in CDNC IDN Table 2015, CGP reviewed the 27 Normalized characters, 145 IICore characters and 7 dotAsia characters, and updated their variant mappings. Among 19738 characters, 3103 of them have non-reflexive variant character(s), including 2214 with 1 variant, 545 with 2 variants, 200 with 3 variants, 89 with 4 variants, 36 with 5 variants, 14 with 6 variants, 4 with 7 variants, 1 with 8 variants. All variant mapping sets are symmetric and transitive.

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.

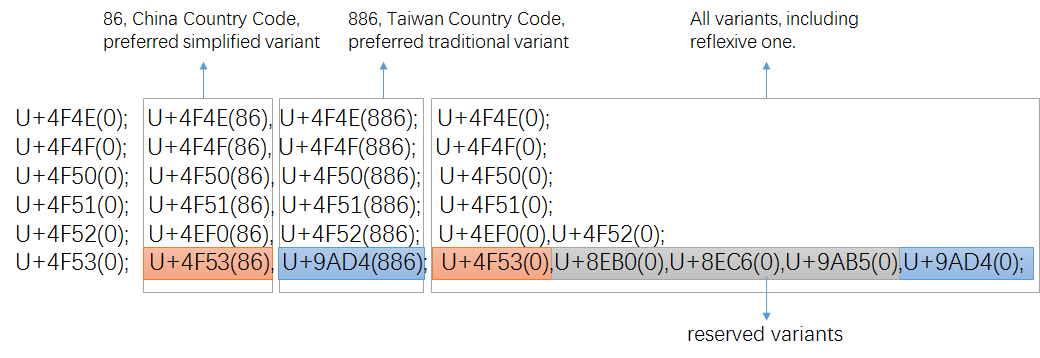


Figure 8: variant setting in CDNC IDN Table

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

|  |  |
| --- | --- |
| “simp” | preferred simplified; |
| “r-simp” | reflexive preferred simplified; |
| “trad” | preferred traditional |
| “r-trad” | reflexive preferred traditional |
| “both” | preferred simplified and traditional ones are the same |
| “r-both” | reflexive preferred simplified and traditional ones are the same |

## 4.2 Coordination between CJK

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 March 2016, KGP published repertoire and variants list Version 0.4, including 4819 Hanja characters with 95 variant characters (47 variant groups). Moreover, KGP raised 260 Hanja/Hanzi character whose variant mappings in CGP LGR are NOT acceptable for KGP.

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 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, 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. Since JGP and KGP has few variant mappings, for any given Chinese character in the overlapping set of CJK repertoire, generally, its CGP variant group is larger than JGP one or KGP one, and contains characters that do not exist in JGP repertoire or KGP repertoire. In that case, JGP and KGP need to import non-repertoire characters from CGP into their repertoires, to meet the requirement “If an LGR includes Han characters, the variant mappings must agree for all three panels.”

Thus, 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 extract its original repertoire with integrated variants from LGR-M.

Step 4: Each CJK GP add “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).

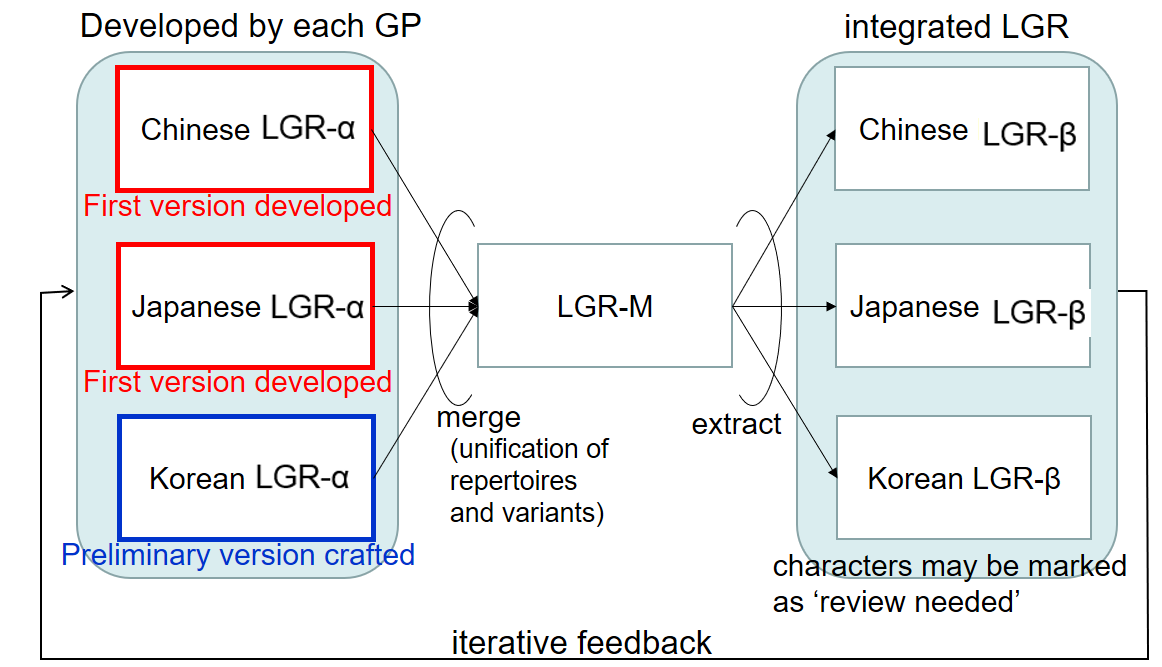


Figure9: Framework of CJK LGR integration for Han characters, by JGP

As introduce in 3.2. CGP repertoire is the superset of JGP repertoire and KGP repertoire, and needs no “out of repertoire” code point.

Another focus problem of code point variant coordination is that not all variant mappings get agreed by all 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 pronounce mean differently in other language environment. 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 260 Hanja/Hanzi character whose variant mappings in CGP LGR are UNACCEPTABLE. As the requesting party for 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 260 character, including the number of registered labels containing disputed character under .CN/.TW/.网址, and the semantics of all these registered labels.

CGP would take different coordination approaches for different situation:

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

This work is still on going and is supposed to be done before ICANN 57.

## 4.3 Recommendation of Variants for TLDs

As above, before the final coordination and the unified variant mapping table are finished, CGP will use the variant mapping in the current work as illustrated in 4.1 and listed in Appendix C.

# Whole Label Evaluation Rules

Following XML-format transforming regulations (Representing Label Gneration Rulesets using XML, <https://datatracker.ietf.org/doc/draft-davies-idntables/>), CGP generates its own XML table of CGP character repertoire, and marks every variant code point with “simp”,”r-simp” “trad”, “r-trad”,“both”,”r-both” or “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 registrant, CGP WLE is designed as follows:

<rules>

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

<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" common="original label"/>

<action disp="blocked" any-variant="simp trad both r-simp r-trad r-both" "block any other mixed labels" />

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

</rules>

The above rules set originally applied label, all-simplified variant labels and all-traditional variant labels ALLOCATALBE, while set any other variant labels BLOCKED.

But still, there is minor difference between CDNC practice and CGP WLE, on the disposition action of “allocatable” label and “blocked” label.

CDNC rules generate many preferred labels, but in practice, CDNC members only delegate the applied label, one preferred SC label and one preferred TC label, no more than 3 labels to the registrants., and reserve the others. Some auxiliary measures are introduced to reduce the number of allocatable labels and related computational complexity, like human interaction process (applicant preference 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 allocable labels, some similar CDNC design should be taken into account, which correspondingly will change WLE rules.

# Possible changes in the future

* Repertoire

Some code points might be removed from the current repertoire according to the coordination results as mentioned in 4.2.

* Variants

Some current variant mapping groups might be decoupled according to the coordination results as mentioned in 4.2.

* Variant Types

To reduce the number of allocatable labels, some new sub-type might be employed as mentioned in chapter 5.

* WLE

New rules are needed if new variant type are employed as mentioned in chapter 5.

# 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 Chinese Characters and Variants Table <http://www.cdnc.org/gb/research/file/unicode.txt>

International Ideographs Core <http://hanzi.unihan.com.cn/downloads/IIcoreMappingTable.txt>

JPRS Japanese Characters Set <http://jprs.jp/doc/rule/saisoku-1-wideusejp-furoku-4.html>

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

DotAsia ZH IDN Language 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 Europe 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.** Prof. She has successfully helped to make structured encoding of Chinese character components which 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 bdfack-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 Hongkong 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：CGP Variants