

**Title: IRGN2343 Proposal to Update PnP to Reflect Existing Practices**

**Date: 2018/06/04**

**Type: Individual Contribution**

**Author: Henry Chan**

## **Introduction**

Some processes specified in the IRG PnP are outdated and do not match the current processes carried out by IRG. This document seeks to amend the IRG PnP such that it is more in-line with current processes, and also suggests minor modifications to the submission.

## **Issue 1: Elaboration of the Non-Cognate Rule**

In actual IRG processes, characters which are unrelated in historical derivation may be unified under exceptional circumstances. This is when the logical structure of the character is identical. For example, 𣎵木几 (table) and 𣎵木几 (c-simplified form of 機) are unified despite being non-cognate.

Suggestion to add the following text to 2.1.3. Non-cognate Rule:

Non-cognate characters which are identical in logical structure may be unified. For example, 𣎵木几 (wooden table) and 𣎵木几 (c-simplified form of 機) are unified despite being unrelated in historical derivation.

Meanwhile, certain characters are consistently disunified by IRG even if the shape is identical, the majority which are characters belonging to the Moon (月) radical and the Meat (肉) radical. Some characters are also related in historical derivation but disunified by IRG.

Suggestion to add the following text to 2.1.3. Non-cognate Rule:

Characters which are systematically distinguished within one or more regions and/or authoritative dictionaries are always disunified even if their shapes may be identical in other regions. For example, characters from Hanzi T sources and Hanzi H sources distinguish characters belonging to the Moon (月) radical and the Meat (肉) radical by varying the direction of the two middle dots in the 月(月) component; the Kangxi dictionary distinguishes between characters belonging to the Moon (月) radical and the Meat (肉) radical by varying the length of the strokes inside the 月(月) component; thus characters that are identical in shape but belong under a different radical (Moon (月) vs Meat (肉)) are disunified. Purely simplified characters which are constrained to the Hanzi (G) source are exempted from this rule.

Characters which are historically derived from each other but used for mutually exclusive reasons may be regarded as non-cognate characters and disunified. For example, 着 is derived from 著 but is used in mutually exclusive meanings and contexts for modern day Hanzi (G and H) and Kanji, are regarded as non-cognate characters and disunified. Submitters are expected to provide proof as outlined in Section 2.4.3.b.

## Issue 2: Naming of Source References

In Section 2.2.1(d)(5), requirements for Source References are as follows:

a) Source Reference to indicate the source and the name of the glyph image for tracking. The source reference should begin with a member body abbreviation (G, H, J, K, KP, M, MY, T, UTC or V) or an international consortium abbreviation (U) followed by no more than 9 characters. It should contain only Latin capital letters and Arabic numbers to indicate the source. Numeric values to indicate the position in a specific source should be followed by a hyphen ("-"). The purpose of source references and an exhaustive list of source references accepted by ISO/IEC 10646 are provided in Section 23 of ISO/IEC 10646. See Annex D for details of member body/international consortium abbreviations.

Since underscores are not allowed for source references in ISO/IEC 10646, the text should be modified as follows (with changes in bold):

a) Source Reference to indicate the source and the name of the glyph image for tracking. The source reference should begin with a member body abbreviation (G, H, J, K, KP, M, MY, T, UTC or V) or an international consortium abbreviation (U) followed by no more than 9 characters. It **must** contain only Latin capital letters and Arabic numbers to indicate the source. Numeric values to indicate the position in a specific source **must** be followed by a hyphen ("-"). **Underscores must not be used.** The purpose of source references and an exhaustive list of source references accepted by ISO/IEC 10646 are provided in Section 23 of ISO/IEC 10646. See Annex D for details of member body/international consortium abbreviations.

### Issue 3: Stroke Count, First Stroke and Total Strokes

Existing text:

- d) Stroke Count of components other than the radical. In case of unified characters, the assignment of stroke count will be based on IRG agreed rules (ref. IRGN954AR and IRGN1105). In case of single source characters, this attribute will be assigned by the IRG Chief Editor following the actual shape.
- e) First Stroke from 1 to 5 as listed in Annex K for components other than the radical. The IRG does not enforce a unified method of first stroke assignment as specified in IRGN954, IRGN954AR, and IRGN1105. It is however advisable for submitters to supply their assignment principles to the technical editor for spotting differences between different submissions. The final decision on first stroke selection rests with the technical editor for IRG editorial work. If the decision of the technical editor is different from that of the submitter, the submitter should be informed of the change.
- j) kTotalStrokes an integer indicating the total number of strokes of a character, including that of its radical. The IRG understands the possible ambiguity of strokes even in individual submissions. This attribute is requested by UTC of Unicode for the Unihan database to maintain locale-related data. The IRG will not check the data and takes no responsibility for its correctness.

In WS2015, IRG agreed to use a unified stroke counting system upon IRGN2221, regardless of the actual glyph shape of submitters. The first stroke would also be reassigned by the Chief Editor in light of the decisions. Furthermore, there was considerable confusion over First Stroke in WS2017.

Therefore, the text should be updated to account for IRGN2221 and the aforementioned decisions. Since a unified stroke counting system is used for stroke count, the first stroke and final stroke count can be trivially derived too.

Lastly, stroke count, first stroke and total strokes should be logically grouped together

- d) Stroke Count of components other than the radical. Assignment of stroke count will be based on IRG agreed rules (ref. IRGN954AR, IRGN1105 and IRGN2221), regardless of the actual shape.
- e) First Stroke of components other than the radical, from 1 to 5 as listed in Annex K. Assignment of stroke count will be based on IRG agreed rules (ref. IRGN954AR, IRGN1105 and IRGN2221), regardless of the actual shape.
- f) Total Stroke Count. An integer indicating the total number of strokes of a character, including that of its radical. Assignment of stroke count will be based on IRG agreed rules (ref. IRGN954AR, IRGN1105 and IRGN2221), regardless of the actual shape.

Per item d), e), and f), the final decision lies in the IRG Technical Editor.

#### **Issue 4: Format of references**

Existing text:

- i) References to evidence documents including document names and other information. Submitters may supply additional information related to evidence submission and additional columns can be added as sub-items of i). For example, in addition to i) as the name of the original evidence document, submitters may include i1) for page number in the document, i2) for row number in that page, i3) to indicate position in that row, and i4) as the name of the JPG file showing the page of an evidence document where the character in question appears.

Should be amended to

- i) References to evidence documents including document names and other information. Submitters may supply additional information related to evidence submission, such as page number, row number and position in the row. For characters with more than one evidence, the information should be separated by semicolon.
- j) Evidence Document(s) File Names. The name of JPG/PDF file(s) to the evidence documents. The file name of each evidence document must be named using the source reference, optionally followed by a "-" (hyphen) and a multi-digit number. For characters with more than one evidence file, the file names should be separated by semicolon.

### Issue 5: File Format of Glyph Image

In section 2.2.1.d.(5) both .bmp and .png are accepted:

- b) Glyph Image File Name. The file name of each glyph image must be named using the source reference followed by “-” and followed by the submission date in the form of “YYMMDD” with a file extension of .bmp in bitmap format or .png in PNG format. Any revision in the future requested either by the submitter or IRG should not change the source reference part, but the new file name should have a new submission date.

The subsequent section 2.2.2a should also be amended:

- a. Glyph Image: Each proposed ideograph must be accompanied by a corresponding 128 x 128 bitmap file in Song or Ming style. The file name should be the same as the source reference (defined in Section 2.2.1.d.(5) with .bmp as its file extension.

To

- a. Glyph Image: Each proposed ideograph must be accompanied by a corresponding 128 x 128 bitmap file / **PNG file** in Song or Ming style. The file name should be the same as the source reference (defined in Section 2.2.1.d.(5) with .bmp **or .png** as its file extension.

#### **Issue 6: 2.3.4. Production of IRG Working Drafts**

The Section 2.3.4. refers to two sets, the M-set (main set) and D-set (Discussion set). However, these are not the general terms used by the current IRG processes.

Also, the IDS checking is currently conducted by Kawabata-san, while the Working Set preparation is prepared by Chief Editor. Given that the IDS checking result is not directly applicable to the initial preparation and must be reviewed, it is better to remove such wording from the PnP.

Existing text:

- a. Division of Character Subsets: By the result of IDS checking, submitted ideographs will be grouped into the following two subsets:
  - (1) M-set (main set): for ideographs with proper IDS and found not to be unifiable with current standardized ideographs or previously discussed ideographs with proper IDS.
  - (2) D-set (discussion set): for ideographs with missing, incomplete, or inconclusive IDS, or ideographs of which the attribute data have been questioned by any member body during a review process, or ideographs that might be unifiable with standardized or previously discussed ideographs. Ideographs with missing or incomplete IDS will be commented as such, and checked intensively through manual checking. Ideographs that might be unifiable with standardized or previously discussed ideographs will also be commented as such, and their suitability for unification must be manually checked. Dis-unification must be supported by evidence.
  
- f. After consolidation, the IRG chief editor and technical editor may ask member bodies to review M-set and D-set based on an agreed IRG review schedule and task division.

Suggested text:

a. Division of Character Subsets: Each Working Set is divided into three subsets:

(1) **M-set** (Main Working Set): all ideographs are initially in this set. Ideographs where any attribute data and/or validity of character and/or evidence is questioned by any member body during each review process is moved to the P-set if not the issue is not resolved during the immediately subsequent IRG meeting. Ideographs which remain the main set at the end will be submitted for encoding.

(2) **U-set** (Unified & Withdrawn Set): characters that are concluded to be unified or withdrawn are moved into the U-set. Given further evidence, a character may be moved back into the M-set.

(2) **P-set** (Pending & Discussion Set): for ideographs where attribute data is missing, inconclusive, or the validity of the character and/or evidence has been questioned by any member body during a review process. Ideographs that might be unifiable with standardized or previously discussed ideographs will also be commented as such, and their suitability for unification must be manually checked. Dis-unification must be supported by evidence. Given further evidence, a character may be moved back into the M-set.

f. After consolidation, the IRG chief editor and technical editor may ask member bodies to review the working set based on an agreed IRG review schedule and task division.

Existing text:

#### 2.4.2. Principles on Manual Checking

- a. Duplication and Unification: For D-set ideographs, member bodies should ensure that they are not duplicates of or unifiable with any ideographs in the standard, working set(s) submitted to WG2, or in the current working set .
- b. Additional Evidence and Arguments: For each proposed ideograph in the D-set that has been questioned for possible unification, the submitter should prepare arguments with further evidence of its use and documentary proof (for example, from dictionaries, legal documents or other publications) showing that it is not unifiable with any standardized ideograph or ideograph proposed in the same or another working draft. (...)
- e. Rejection: Questioned ideographs with no counter arguments in support of dis-unification supplied to the IRG meeting will be automatically marked as unified.

Suggested changes:

#### 2.4.2. Principles on Manual Checking

- a. Duplication and Unification: For **M-set ideographs**, member bodies should ensure that they are not duplicates of or unifiable with any ideographs in the standard, working set(s) submitted to WG2, or in the current working set.
- b. Additional Evidence and Arguments: For each proposed ideograph in the **M-set** that has been questioned for possible unification, the submitter should prepare arguments with further evidence of its use and documentary proof (for example, from dictionaries, legal documents or other publications) showing that it is not unifiable with any standardized ideograph or ideograph proposed in the same or another working draft. (...)
- e. Rejection: Questioned ideographs with no counter arguments in support of dis-unification supplied to the IRG meeting will be automatically marked as unified **and moved to the U-set**.

In existing processes, characters with changes in radical decided during meetings are not moved to discussion nor reviewed separately from other existing characters in the main set. Suggest removing the following text:

When the Kangxi radical or stroke count of an ideograph is found to be incorrect, the ideograph will be moved to D-set pending another manual review to prevent any unification errors caused by not having conducted the review with ideographs having the correct Kangxi radical or stroke count.



Suggest adding the following text to the end of 2.4.3. Submission of Possibly Unifiable Ideographs:

- g. **Unresolved issues:** ideographs with unresolved issues are moved to the P-set.

Suggest correcting the footnote on page 12 per Section 3.1.a.:

<sup>1</sup> The first IRG collection using this naming convention is IRG Collection 2015.

To

<sup>1</sup> The first IRG collection using this naming convention is IRG **Working Set** 2015.

Suggest amending the following text:

3.2.b. After serial numbers have been assigned, submitted ideographs must undergo IDS checking to detect any duplication and unification. Based on the result of IDS checking as described in Section 2.3.3, submitted ideographs will be grouped into M-set and D-set as described in Section 2.3.4.

3.2.c. After consolidation, the working drafts will be assigned an IRG document number with a version number. They will be distributed to member bodies' editors and made available on the official IRG website so that any other experts can have access to them. The IRG chief editor and technical editor may assign member bodies' editors to check M-set and D-set ideographs for either the entire working set or certain portions of it depending on their reasonable estimation of the workload.

To

3.2.b. After serial numbers have been assigned, the working drafts will be assigned an IRG document number with a version number. They will be distributed to member bodies' editors and made available on the official IRG website so that any other experts can have access to them. The IRG chief editor and technical editor may assign member bodies' editors to check either the entire working set or certain portions of it depending on their reasonable estimation of the workload.

3.2.c. The IDS checker will conduct IDS checking on the working set to detect any duplication and unification. Result of IDS checking will be discussed in the subsequent meeting.

Suggest amending the following text:

3.3.a. Each member body's editor must check the assigned M-set and D-set for data integrity, correctness, missing data and duplication. Checking for unification is not mandatory, but desirable. Typical review comment examples for each set are provided in Section 4.

3.4. b. All the conclusions must be agreed and endorsed by the IRG plenary in its resolutions. As a result of the resolutions, some ideographs may be removed or moved between M-set and D-Set.

3.4. c. The IRG technical editor will create a new M-set and D-set one month after the IRG meeting, and register them as IRG documents with version information.

3.5. a. Each member body's editor must check the newly created M-set and D-set for correctness and duplication.

3.6. b. All the conclusions must be agreed and endorsed by the IRG plenary in its resolutions. As a result of the resolutions, some ideographs may be removed or moved between M-set and D-set.

3.6. c. The IRG technical editor will create a new M-set and D-set one month after the IRG meeting, and produce a registered IRG document.

3.7.a. All member bodies' editors are requested to check M-set intensively based on comments and conclusions made at all previous stages. At the final checking stage, no ideographs are allowed to be moved from D-Set to M-Set although ideographs in the M-set can still be moved to D-set if problems are found.

3.8.b. If there is no positive decision on an M-set ideograph, it will be moved to D-set. Again, no character will be moved from D-set to M-set at this stage. Ideographs may only be moved from M-set to D-set if problems are found.

Once M-set is completed for submission to WG2, records of characters in the D-set will no longer be maintained by the IRG. Characters remained in the D-set can be re-submitted in future extensions if pending problems are solved.

4.1. The ultimate target of M-set is a standardized ideograph set. As such, it must be carefully examined. If any suspicious characters are found, they will be moved to D-set or removed from the current working set altogether.

To

3.3.a. Each member body's editor must check the assigned Working Set for data integrity, correctness, missing data and duplication. Typical review comment examples for each set are provided in Section 4.

3.4.b. All the conclusions must be agreed and endorsed by the IRG plenary in its resolutions. As a result of the resolutions, some ideographs may be removed or moved between M-set, U-Set and P-Set.

3.4.c. The IRG technical editor will prepare a new version of the Working Set one month after the IRG meeting and register them as IRG documents with version information.

3.5.a. Each member body's editor must check the new version of the Working Set for correctness and duplication.

3.6.b. All the conclusions must be agreed and endorsed by the IRG plenary in its resolutions. As a result of the resolutions, some ideographs may be removed or moved between M-set, U-set and P-set.

3.6.c. The IRG technical editor will prepare a new version of the Working Set one month after the IRG meeting, and produce a registered IRG document.

3.7.a. All member bodies' editors are requested to check M-set intensively based on comments and conclusions made at all previous stages. At the final checking stage, no ideographs are allowed to be moved from P-Set to M-Set although ideographs in the M-set can still be moved to P-set if problems are found.

3.8.b. If there is no positive decision on an M-set ideograph, it will be moved to P-set. Again, no character will be moved from P-set to M-set at this stage. Ideographs may only be moved from M-set to P-set if problems are found.

Once M-set is completed for submission to WG2, records of characters in the P-set and U-set will no longer be maintained by the IRG. Characters remained in the P-set can be re-submitted in future extensions if pending problems are solved.

4.1. The ultimate target of M-set is a standardized ideograph set. As such, it must be carefully examined. If any suspicious characters are found, they will be moved to P-set or U-set.

We should also explicitly distinguish the cases between unification and IVS.

Modify:

Possible Comment by a Reviewer	Possible Resolution
Wrong or Missing Glyph	<ul style="list-style-type: none"><li>● The wrong glyph is corrected, or the missing glyph supplied. The ideograph will be moved to D-set for manual checking.</li></ul>
Wrong Kangxi radical / stroke count / first stroke	<ul style="list-style-type: none"><li>● Data will be corrected and the ideograph will be moved to D-set for further manual checking.</li></ul>
Wrong IDS	<ul style="list-style-type: none"><li>● IDS will be corrected and the character will be moved to D-set for checking by the IDS checker.</li><li>● Move to D-set (in case IDS cannot be corrected).</li></ul>
May be unifiable with U+xxxxx (standardized ideograph)	<ul style="list-style-type: none"><li>● Unified with U+xxxxx and the submitter will request a new source reference to U+xxxxx.</li></ul>

	<ul style="list-style-type: none"> <li>● Unified with U+xxxxx and the submitter will request that this character be treated as a Compatibility Ideograph.</li> <li>● Unified to U+xxxxx and this entry will be removed. (May consider to register it as IVS.)</li> <li>● Not unifiable.</li> </ul>
May be unifiable with xxxxx (M-set ideograph)	<ul style="list-style-type: none"> <li>● Unified with xxxxx and this source reference will be attached to xxxxx.</li> <li>● Unified with xxxxx and the submitter may consider registering it as a Compatibility Ideograph Character or IVS.</li> <li>● Not unifiable.</li> </ul>

To

Possible Comment by a Reviewer	Possible Resolution
Wrong or Missing Glyph	<ul style="list-style-type: none"> <li>● The wrong glyph is corrected, or the missing glyph supplied.</li> </ul>
Wrong Kangxi radical / stroke count / first stroke	<ul style="list-style-type: none"> <li>● Data will be corrected.</li> </ul>
Wrong IDS	<ul style="list-style-type: none"> <li>● IDS will be corrected</li> </ul>
May be unifiable with U+xxxxx (standardized ideograph)	<ul style="list-style-type: none"> <li>● Unify to U+xxxxx and the submitter will request a horizontal extension to U+xxxxx.</li> <li>● Unify to U+xxxxx and entry will be removed.</li> <li>● Withdraw for registration as an ideographic variant via the IVD to U+xxxxx</li> <li>● Not unified</li> </ul>
May be unifiable with xxxxx (M-set ideograph)	<ul style="list-style-type: none"> <li>● Unified with xxxxx and the source reference will be added to xxxxx.</li> <li>● Unified with xxxxx and the entry is dropped.</li> <li>● Withdrawn for registration as an ideographic variant via the IVD to xxxxx.</li> <li>● Not unified.</li> </ul>

## Issue 7: IDS characters

The current characters allowed by the IDS are specified in Annex B which points to IRGN1183. In IRNG1183, allowed characters to be CDC are specified as follows:

CDC (Character description component): A UCS character that is included either in CJK UNIFIED IDEOGRAPHS, in CJK UNIFIED IDEOGRAPHS EXTENSION A, in CJK UNIFIED IDEOGRAPHS EXTENSION B, in KANGXI RADICALS, in CJK RADICALS SUPPLEMENT, or in CJK COMPATIBILITY IDEOGRAPHS. In other words, CDC is a DC that consists of just one UCS character.

Annex B of IRG PnP also states:

Each member body should consult IRGN1183 on IDS. In addition to the Character Description Components (CDC) defined in IRGN1183, all CJK Unified Ideographs accepted by ISO/IEC 10646 in its amendments are also qualified as CDC in constructing IDS.

First, since IRG now adopts a unified stroke counting method, the stroke count and total stroke count can be derived and checked from the IDS data. For highly deformed characters, it is useful to use ① ~ ⑳ to represent unencoded components by their stroke count such that the stroke count and total stroke count can be automatically derived and checked.

The majority of CJK Compatibility Ideographs are merely glyphic or pronunciation variants of existing unified characters. Kangxi Radicals are also identical to existing CJK Unified Ideographs. Both types of characters are normalized to their unified equivalents in Unicode-compliant text processing. To avoid complicating the IDS algorithm, Kangxi Radicals and Compatibility Ideographs should be forbidden.

There are some characters in the CJK Radicals Supplement which are highly similar or identical to existing CJK Unified Ideographs. They should also be forbidden with certain exceptions. Suggested list as follows:

Component	Codepoint	Restriction	
ㄨ	<b>2E80</b>	<b>Allowed</b>	
ㄐ	2E81	Forbidden, use ㄐ	20086
ㄑ	2E82	Forbidden, use ㄑ	4E5B
ㄒ	2E83	Forbidden, use ㄒ	4E5A
ㄓ	<b>2E84</b>	<b>Allowed</b>	
ㄔ	2E85	Forbidden, use ㄔ	4EBB
ㄕ	<b>2E86</b>	<b>Allowed</b>	
ㄖ	<b>2E87</b>	<b>Allowed</b>	
ㄗ	2E88	Forbidden, use ㄗ	2008A
ㄘ	2E89	Forbidden, use ㄘ	5202
ㄙ	<b>2E8A</b>	<b>Allowed</b>	
ㄚ	2E8B	Forbidden, use ㄚ	353E
ㄛ	2E8C	Allowed	
ㄜ	2E8D	Forbidden, use ㄜ	2D544

兀	2E8E	Forbidden, use 兀	5140
允	2E8F	Forbidden, use 允	5C23
尢	2E90	Forbidden, use 尢	5C22
允	2E91	Forbidden, use 允	5C23
巳	2E92	Forbidden, use 巳	5DF3
么	2E93	Forbidden, use 么	5E7A
彑	2E94	Forbidden, use 彑	5F51
𠄎	2E95	Forbidden, use 𠄎	5F50
𠄎	2E96	Forbidden, use 𠄎	5FC4
小	2E97	Forbidden, use 小	38FA
扌	2E98	Forbidden, use 扌	624C
夂	2E99	Forbidden, use 夂	2E99
无	2E9B	Forbidden, use 无	65E1
𠄎	2E9C	Forbidden, use 𠄎	5183
月	2E9D	Forbidden, use 月 or 月	6708/2EBC
𠄎	2E9E	Forbidden, use 𠄎	6B7A
母	6BCD	Forbidden, use 母	6BCD
民	2EA0	Forbidden, use 民	6C11
彡	2EA1	Forbidden, use 彡	6C35
氷	2EA2	Forbidden, use 氷	6C3A
灬	2EA3	Forbidden, use 灬	706C
𠄎	2EA4	Forbidden, use 𠄎	722B
𠄎	2EA5	Forbidden, use 𠄎	722B
𠄎	2EA6	Forbidden, use 𠄎	4E2C
𠄎	<b>2EA7</b>	<b>Allowed</b>	
彡	2EA8	Forbidden, use 彡	72AD
𠄎	2EA9	Forbidden, use 𠄎	248E9
𠄎	2EAA	Forbidden, use 𠄎	24D14
𠄎	2EAB	Forbidden, use 𠄎	7F52
𠄎	2EAC	Forbidden, use 𠄎	793A
𠄎	2EAD	Forbidden, use 𠄎	793B
𠄎	2EAE	Forbidden, use 竹	7AF9
糸	2EAF	Forbidden, use 糸	7CF9
𠄎	2EB0	Forbidden, use 𠄎	7E9F
𠄎	2EB1	Forbidden, use 𠄎	7F53

𠂇	2EB2	Forbidden, use 𠂇	7F52
𠂈	2EB3	Forbidden, use 𠂈	34C1
𠂉	2EB4	Forbidden, use 𠂈	34C1
𠂊	2EB5	Forbidden, use 𠂊	2626B
𠂋	<b>2EB6</b>	<b>Allowed</b>	
𠂌	<b>2EB7</b>	<b>Allowed</b>	
𠂍	2EB8	Forbidden, use 羊	7F8A
𠂎	2EB9	Forbidden, use 𠂎	8002
𠂏	2EBA	Forbidden, use 𠂏	8080
𠂐	<b>2EBB</b>	<b>Allowed</b>	
月	<b>2EBC</b>	<b>Allowed (use 月 or 月 for moon vs meat radical)</b>	<b>6708 or 2EBC</b>
白	2EBD	Forbidden, use 白 or 白	81FC or 26951
𠂑	2EBE	Forbidden, use 𠂑	8279
𠂒	2EBF	Forbidden, use 𠂑	8279
𠂓	<b>2EC0</b>	<b>Allowed (use 𠂑 or 𠂓 for grass vs horn)</b>	<b>8279 or 2EC0</b>
虎	2EC1	Forbidden, use 虎	864E
𠂔	2EC2	Forbidden, use 𠂔	8864
𠂕	2EC3	Forbidden, use 𠂕	8980
西	2EC4	Forbidden, use 西	897F
见	2EC5	Forbidden, use 见	89C1
角	2EC6	Forbidden, use 角	89D2
𠂖	2EC7	Forbidden, use 角 or 𠂖	89D2 or 278B2
讠	2EC8	Forbidden, use 讠	8BA0
贝	2EC9	Forbidden, use 贝	8D1D
𠂗	2ECA	Forbidden, use 𠂗	27FB7
车	2ECB	Forbidden, use 车	8F66
辶	2ECC	Forbidden, use 辶	8FB6
辶	2ECD	Forbidden, use 辶	8FB6
辶	2ECE	Forbidden, use 辶	8FB6
𠂘	2ECF	Forbidden, use 𠂘	961D
𠂙	2ED0	Forbidden, use 𠂙	9485
長	2ED1	Forbidden, use 長	9577
𠂚	2ED2	Forbidden, use 𠂚	9578
长	2ED3	Forbidden, use 长	957F
门	2ED4	Forbidden, use 门	95E8

𠄎	2ED5	Forbidden, use 𠄎	28E0F
𠄐	2ED6	Forbidden, use 𠄐	961D
𠄑	2ED7	Forbidden, use 雨	96E8
青	2ED8	Forbidden, use 青	9752
韦	2ED9	Forbidden, use 韦	97E6
页	2EDA	Forbidden, use 页	9875
风	2EDB	Forbidden, use 风	98CE
飞	2EDC	Forbidden, use 飞	98DE
食	2EDD	Forbidden, use 食	98DF
𠄒	2EDE	Forbidden, use 𠄒	2967F
𠄓	2EDF	Forbidden, use 𠄓	98E0
𠄔	2EE0	Forbidden, use 𠄔	9963
𠄕	2EE1	Forbidden, use 𠄕	29810
马	2EE2	Forbidden, use 马	9A6C
骨	2EE3	Forbidden, use 骨	9AA8
鬼	2EE4	Forbidden, use 鬼	9B3C
鱼	2EE5	Forbidden, use 鱼	9C7C
鸟	2EE6	Forbidden, use 鸟	9E1F
𠄖	2EE7	Forbidden, use 𠄖	5364
麦	2EE8	Forbidden, use 麦	9EA6
黄	2EE9	Forbidden, use 黄	9EC4
𠄗	2EEA	Forbidden, use 𠄗	9EFE
齐	2EEB	Forbidden, use 齐	6589
齐	2EEC	Forbidden, use 齐	9F50
𠄘	2EED	Forbidden, use 𠄘	6B6F
𠄙	2EEE	Forbidden, use 𠄙	9F7F
𠄚	2EEF	Forbidden, use 𠄚	7ADC
龙	2EF0	Forbidden, use 龙	9F99
龜	2EF1	Forbidden, use 龜	9F9C
龜	2EF2	Forbidden, use 龜	4E80
龟	2EF3	Forbidden, use 龟	9F9F



Lastly, there are some characters which differ only in the slant of the last stroke (横 vs 提). The flat version should be used for consistency. Some other positional variants were encoded as ideographs. Also, for consistency with the IDS maintained by KAWABATA Taichi for IRG use, an additional pair for 竹 is included. Table as follows:

Non-preferred Character		Recommended Character	
𠄎	248E9	王	738B
孑	5B51	子	5B50
𠄎	725C	牛	725B
卓	9FBA	卓	2099D
繚	9FBB	繚	470C
𠄎	25AD7	竹	7AF9

Some characters are disunified per Source Code Separation, and their stroke count and/or first stroke differ. Per the unified stroke counting rule and first stroke conventions, the Kangxi form is generally preferred. List as follows

Prohibited Character		Target Character		Reason
争	4E89	爭	722D	SC different
𠄎	5942	𠄎	5950	SC different
戶	6236	户	6237	FS different
戶	6238	户	6237	FS different
既	65E2	既	65E3	SC/FS different
冊	518C	冊	518A	FS different
并	5E76	并	5E77	SC/FS different
虛	865A	虛	865B	SC different

For duplicate characters, the smaller code point is preferred. List as follows:

Prohibited	Preferred
𠄎 (U+3BA3)	𠄎 (U+69E9)