INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 2/WG 2/IRG

Universal Coded Character Set
(UCS)

ISO/IEC JTC 1/SC 2/WG 2/IRG *N1920Draft-3*
(Revision of IRGN1772/IRGN1823)
2013-05-04

Title:	**IRG Principles and Procedures Version 6**
Source:	**IRG P&P Drafting Group**
Action:	**For review by the IRG and WG2**
Distribution:	**IRG Member Bodies and Ideographic Experts**
Editor in chief:	**Lu Qin, IRG Rapporteur**
References:	**IRG 1920Draft(2012-11-15) and Japan's feedback(2013-04-23), ROK feedback(2013-05-16 and 2013-05-21) and IRG Meeting No. 40 discussions**
	IRG 1823 Draft 3 and feedback from HKSARG, Korea and IRG Meeting No. 39 discussions
	IRGN1823 Draft2 feedback from HKSARG and Japan
	IRG N1823Draft_gimgs2_Feedback
	IRG N1781 and N1782 Feedback from KIM Kyongsok
	IRGN1772 (P&P Version 5)
	IRG N1646 (P&P Version 4 draft)
	IRG N1602 (P&P Draft 4) and IRG N1633 (P&P Editorial Report)
	IRG N1601 (P&P Draft 3 Feedback from HKSAR)
	IRG N1590 and IRGN 1601(P&P V2 and V3 draft and all feedback)
	IRG N1562 (P&P V3 Draft 1 and Feedback from HKSAR)
	IRG N1561 (P&P V2 and all feedback)
	IRG N1559 (P&P V2 Draft and all feedback)
	IRG N1516 (P&P V1 Feedback from HKSAR)
	IRG N1489 (P&P V1 Feedback from Taichi Kawabata)
	IRG N1487 (P&P V1 Feedback from HKSAR)
	IRG N1465, IRG N1498 and IRG N1503 (P&P V1 drafts)

Table of Contents

1. Introduction

This document is a standing document of the ISO/IEC JTC 1/SC 2/WG 2/IRG for the standardization of Chinese-Japanese-Korean (CJK) Unified Ideographs.   It consists of a set of principles and procedures on a number of items relevant to the preparation, submission and development of repertoires of CJK Unified Ideographs extensions for additions to the standard (ISO/IEC 10646).   Submitters should check the standard documents (including all the amendments and corrigenda) before preparing new submissions.

For any issue that is not explicitly covered in this document, the IRG will follow the Principles and Procedures of WG2 and other higher level directives.

1.1. Scope of IRG Work

The IRG works on CJK ideograph-related tasks under the supervision of WG2 (SC2 Resolution M17-08). The following is a list of current and completed IRG projects:

   a.   CJK Unified Ideograph Repertoire and its extensions
   b.   Kangxi Radicals and CJK Radical Supplements
   c.   Ideographic Description Characters
   d.   International Ideographs Core (IICore)
   e.   CJK Strokes

Work on new IRG projects requires the approval of WG2 and preparation of relevant documents for such approval is required before the IRG can officially launch new projects.

1.2. Scope of This Document

The following sections are dedicated to the standardization of CJK Unified Ideographs, describing the set of principles and procedures to be applied in the development of a new repertoire of CJK Unified Ideographs as specified under work item a. in Section 1.1. In addition, it specifies the maintaianance information of IRG website and document registration procedure of IRG in Section 5 and 6, respectively.

This document does not cover other IRG work items listed in Section 1.1.   Standardizing CJK Compatibility ideographs maintained in UCS for the purpose of round-trip integrity with other standards is out of the IRG scope. However, CJK compatibility characters submitted to WG2 must be reviewed by the IRG to avoid potential problems. For the handling mis-unification and duplicate ideographs, WG2 Principles and Procedures Annex I and J attached to this document should be referenced.

2. Development of CJK Unified Ideographs

Any new extension work must be approved by WG2 before the actual consolidation and review can be formally carried out. There are no fixed rules to initiate a new extension. Normally, some member bodies would initiate it by submitting a proposal which states the need of a required repertoire. Submission of proposals must follow the principles and procedures stated in this document. The IRG would first review the proposal and confirm that it is within the IRG scope.

Taking into considerations of (1) the urgency, (2) the justification and the repertoire size of the proposal, and (3) the current workload of the IRG, the IRG may take one of the following actions.
   a.   Endorse the proposal and submit it to WG2 for approval.
   b.   Request other member bodies to submit characters of similar nature so as to estimate the real workload before submitting the proposal to WG2 for endorsement.
   c.   Accept the proposal as a contribution to an ongoing IRG work item.
   d.   Reject the proposal with justifications. A rejected proposal may be revised and re-submitted to the IRG.

2.1. Principles on Identification of CJK Unified Ideographs

2.1.1. Principles on Encoding

Ideographs that have the same abstract shapes are unified under the unification rule (Annex S of ISO/IEC 10646) and assigned a single character code. A CJK ideographic character can be represented in many actual forms depending on the writing style adopted. Examples of common writing styles include Song style and Ming style as typical print forms, Kai style as a hand written form, and Cao style as a cursive form. Stylistically different forms of the same character can be represented by a different number or different type of strokes or components, which may affect identification of the same abstract shape. In order to reach a common ground to identify those abstract shapes to be encoded as distinct CJK Unified Ideographs, the IRG only accepts submissions using a print form of glyphs (usually Song style or Ming style).

2.1.2. Unification Procedures of CJK Ideographs

Standard print forms of CJK ideographs are constructed with a combination of known components or stroke types. Many are determined by two components - a radical chosen to classify the character in dictionaries and possibly reflect the meaning of the character, and a phonetic component which represents the pronunciation of the character. Basically, two submitted print forms of glyphs with different radicals are distinct characters even if they they may be considered ideograph variant forms and have the same phonetic component such as '理'(U+7406) and '鯉'(U+9BC9). For non trivial cases, further shape analysis must be conducted. Two similar glyphs are decomposed into radicals, components or stroke types and evaluated by following the unification procedures described in Annex S of ISO/IEC 10646.

2.1.3. Non-cognate Rule

Ideographs that are unrelated in historical derivation (non-cognate characters) are not unified. The following gives examples of semantically different characters with very similar glyphs, they are considered to have different abstract shapes because they are non-cognate.

'戌'(U+620C) and '戍'(U+620D) differ only in rotated strokes or dots (S.1.5 a).
'于'(U+4E8E) and '干'(U+5E72) differ only in folding back at the stroke termination (S.1.5 f).

Because shape analysis alone may not tell non-cognateness or semantic differences, it is the submitter's responsibility to provide information and supporting evidence in order to invoke the non-cognate rule.

2.1.4. Maintaining Up-to-date Unification/Non-unification Examples

In Annex S, unification/non-unification examples are summarized from past practice and they are not exhaustive. If there is ambiguity in applying these rules, the IRG must first have a formal discussion for agreement. In case of finding worthy examples for recording, the IRG will maintain an up-to-date list of unification/non-unification examples by adding such examples in IRG working document series(IWDS) maintained at the IRG website. Furthermore, the list will be reported to WG2 from time to time as an input for Annex S revision. Detailed procedures for IWDS update is given in Annex E.

2.2. Principles on Submission of Ideographs to the IRG

2.2.1. Basic Rules on Submission and Required Data to be Submitted

A member body may submit the following to the IRG along with its repertoire. Submission may be categorized by the content of the proposal as specified below.

a. **New Sources to Standardized Ideographs**. If the submission specifies new sources (such as an existing or a new national standard) to some existing standardized Ideographs, these new sources need to be reviewed and approved by the IRG before submission to WG2. Sources and source references in the current ISO/IEC 10646 standard can be found in clause 23 of ISO/IEC 10646 Third Edition (2012-06-01)(See Annex D for an up-to-date IRG list of sources).

b. **New Sources to Working Sets**. In case there are some remaining characters with new sources in previous standardization stages, new sources reviewed and approved by the IRG

will be incorporated by the IRG technical editor into the up-to-date IRG list of sources for the current IRG working sets.

c.   **New CJK Compatibility Ideographs (Vertical extension).** In case a member body needs to add CJK Compatibility Ideographs, the member body needs to supply the following information which will be reviewed by the IRG before submission to WG2 to avoid possible problems of unification or dis-unification with other CJK Unified Ideographs.

(1) Table showing the following data for each CJK Compatibility Ideograph
   a) Code position of the corresponding UCS CJK Unified Ideograph
   b) Glyph(s) of the corresponding UCS CJK Unified Ideograph
   c) Glyph of the CJK Compatibility Ideograph to be printed in the member body's column of CJK Compatibility Ideographs Code Table
   d) Source reference (for detailed format, see 2.2.1.d.(5).a)
   e) evidence showing why the CJK Compatibility Ideograph needs to be added to UCS (e.g., a national standard showing two distinct code positions for two glyphs that are one and the same).
(2) TrueType font containing glyphs to be printed in the member body's column of CJK Compatibility Ideographs Code Table (for detailed format, see 2.2.2.b.)

d.   **New CJK Unified Ideographs (Vertical extension).**   All CJK Unified Ideograph submissions are subject to the following rules:

(1). **Collection Size for the Working Set**: As the collection is defined by submitters according to their own criteria, the IRG will not impose a limit on the collection size.   However, to rationalize the burden of the checking process and to achieve a high quality of standard within a reasonably short period of time, the size of the collection to be reviewed by IRG member bodies as the current working set normally cannot exceed 4,000 ideographs. Based on this principle, member bodies may be asked to divide their submitted collections into subsets to be processed in different IRG collections. The guideline for forming the current working set is given in Annex

(2). **Pre-submission Unification Checking**: A member body should be EXTREMELY CAREFUL **not to submit CJK Unified Ideographs that are already standardized or previously discussed** and recorded at IRG meetings. By the nature of ideographs, it is very difficult for IRG reviewers to find out all unifiable ideographs. Thus, it is important to achieve high quality at the time of submission. Submitters must make sure that submitted ideographs do not fall into any of the following categories:
   a)   Ideographs already standardized in the ISO/IEC 10646 standard (including amendments).
   b)   Ideographs currently in WG2's working drafts.
   c)   Ideographs currently in IRG working sets including both M-sets and D-sets.
   d)   Ideographs mis-unified or over-unified with ideographs in the current standard based on the list maintained by the IRG in its working document series, IWDS_MUI and IWDS_NUC, respectively.

Low quality submissions may be rejected by applying the "5% Rule" described in Section 2.2.5 below.

(3). **Document Registration**: All submission documents should be registered as IRG documents with an IRG document number(IRGN), whose file name should be in the form of:

IRGN*nnnn_mmmm_ppp_*sub_*sss*

where *nnnn* indicates an IRG document number assigned by the IRG Rapporteur, *mmmm* indicates the member body's abbreviation (as listed in 2.2.1.d.(5).a), *ppp* indicates the working set or repertoire name (such as Ext. X labelled by "_X"), *sss* can be any member body designated indicator, and "sub" is a short hand for submission.

(4). **Submission of Over-unified or Mis-unified Ideographs**: Submission of ideographs that are already mis-unified or over-unified within the current standard should follow the principles in Annex I of WG2 Principles and Procedures. The list of over-unified or mis-unified ideographs should be maintained by the IRG technical editor and made available for update in the IRG working document series IWDS_NUC and IWDS_MUI according the maintenance procedure defined in Annex E of this document. For

mis-unfied non-cognate characters, request can be made to add new code point for disunified characters.

(5). The following data items for each proposed ideograph must be submitted with CSV (Comma Separated Value) text format (in UTF-8) or Microsoft Excel format file:

* Sequence number starting from 1 is required in the first column of each row.
   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, T, H, M, J, K, KP, MY, U or V)[1] or Z for international consortiums followed by no more than 9 characters and should contain only Latin capital letters, Arabic numbers, and hyphens. The purpose of source references and accepted source references by ISO/IEC 10646 are exhaustively listed in Section 23 of ISO/IEC 10646. See Annex D for details on information about member body abbreviations.
   b) *Glyph Image file name.* The file name of each glyph image must be the same as the source reference with file extension of .bmp in bitmap format or .PNG format.
   c) *KangXi Radical Code* from 1 to 214[2] with an additional 0 or 1 to indicate a traditional radical(0) or simplified radical(1) for the glyph shapes given in Annex A.a.
   d) *Stroke Count* of the non-radical component. In case of unified characters, the assignment of stroke count will be based on IRG agreed rules in ref. IRGN954AR and IRGN1105.
   e) **First Stroke** from 1 to 5 listed in Annex K for the remaining components other than the radical. When submitters supply the first stroke information in its submissions, the IRG does not enforce a unified method of first stroke assignment (IRGN954, IRGN954AR, and IRGN1105). It is advisable to supply member body's assignment principles to the technical editor for spotting differences by different submissions. The final decision on first stroke selection rest with the technical editor for IRG editorial work. If the decision of the technical editor is different from that of the member body, the member body should be informed of the change.
   f) *Flag to show whether the ideograph is traditional (0) or simplified (1).*
   g) *Ideographic Description Sequence(IDS)* (ref. IRGN1183).
   h) *Similar Ideographs* if available (identified by their code points in the standard in the form of U+xxxxx) or enter "No" if no known variants, leave it empty if not checked.
   i) *References to evidence documents* including evidence document name and other information. Members can supply additional information related to evidence submissions 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, members can 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.
   j) *Optional information* in text format can be included as additional columns starting from j) .

Some sample submissions are listed in Annex G for reference.
A blank form in Excel format is available for member bodies use as a separate document.

e. **Existing CJK Compatibility Ideographs (Horizontal extension).** In case a member body needs to add new source references to existing CJK Compatibility Ideographs, the member body needs to supply the following information which will be reviewed by the IRG before submission to WG2 to avoid possible problems.

   (1) Table showing the following data for each CJK Compatibility Ideograph
       a) Code position of the existing UCS CJK Compatibility Ideograph
       b) Glyph of the existing UCS CJK Compatibility Ideograph

---

[1] Member body abbreviations correspond to the source standard categories in ISO/IEC 10646 Section 23 except *MY*.

[2] The corresponding code range for Kangxi radicals in ISO/IEC 10646 is from U+2F00 to U+2FD5

| Deleted: , respectively |
| Deleted: . |
| Deleted: (|
| Deleted:) |
| Deleted: *and Variant Ideographs* |

c) Code position of the corresponding UCS CJK Unified Ideograph

d) Glyph(s) of the corresponding UCS CJK Unified Ideograph

e) Glyph of the Compatibility Ideograph in the source reference

f) Glyph of the Compatibility Ideograph to be printed in the member body's column of CJK Compatibility Ideographs Code Table

g) Source reference (for detailed format, see 2.2.1.d.(5).a)

h) evidence showing why a new source reference for the CJK Compatibility Ideograph needs to be added to UCS (e.g., a national standard showing two distinct code positions for two glyphs that are one and the same)

(2) TrueType font containing glyphs to be printed in the member body's column of CJK Compatibility Ideographs Code Table (for detailed format, see 2.2.2.b.)

f. Existing **CJK** Unified Ideographs (Horizontal extension). In case a member body needs to add new source references to existing CJK Unified Ideographs, the member body needs to supply the following information. These characters must be reviewed by the IRG before submission to WG2 to avoid possible problems.

(1) Table showing the following data for each CJK Unified Ideograph

a) Code position of the existing UCS CJK Unified Ideograph

b) Glyph of the existing UCS CJK Unified Ideograph

c) Glyph of the CJK Unified Ideograph to be printed in the member body's column of CJK Unified Ideographs Code Table

d) Source reference (for detailed format, see   2.2.1.d.(5).a)

e) evidence showing why a new source reference for the CJK Unified Ideograph needs to be added to UCS (e.g., a national standard showing the relevant glyph)

(2) TrueType font containing glyphs to be printed in the member body's column of CJK Unified Ideographs Code Table (for detailed format, see 2.2.2.b.)

2.2.2. Required Font to be Submitted

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 below in Section 2.2.1.d.(5).a) with .bmp as its file extension.

b. **TrueType Font** (optional): TrueType font availability is highly recommended although not necessary. Font specification can be found under point 5 of A.1. – Submitter's Responsibilities in Annex A of the Principles and Procedures for UCS provided by WG2[3]. The IRG at certain stage of project development will set a deadline for TrueType font submission.

2.2.3. Required Evidence to be Submitted

a. **Supporting Evidence**: Evidence of the proposed glyph shape, its usage and context with pronunciations, meanings, etc.should be supplied to convince the IRG that it is actually used or non-cognate with other similar ideographs. Evidence for each character must be supplied as scanned images. The provision of evidence on character usage including those for personal names should not be exempted. A declaration for character use without accompanying evidence is generally not acceptable. Considering privacy issues, the IRG has suggested some compromised provisions. Details are given in Annex G Part 3.
Note: To support e-government related initiatives, IRG can at its discretion accept submissions of characters that are used in computer systems administered by government bodies for public service with wide access by government agencies and citizens. Factors considered for such acceptance are further elaborated in Annex G.4.

b. **Questionable Characters** (optional): For candidate ideographs with possible unification questions, submitters are encouraged to provide detailed evidence of use from authoritative sources, and relationships to other standardized ideographs or variants having similar shape or

Deleted: f

Deleted: g

meaning encoded in UCS for review. Characters that have provided this information is not counted as problem character for quality assurance assessment given in 2.2.5.

c. **Avoidance of Derived Simplified Ideographs:** To avoid encoding derived simplified characters that are not in actual use, submission of simplified ideographs requires the actual usage evidence. Providing only their corresponding traditional ideographs will not be considered evidence.

2.2.4. Required Summary Form to be Submitted

Each submission for an ideograph collection should be accompanied by a duly completed "Proposal Summary Form for Additions of CJK Unified Ideographs to the Repertoire of ISO/IEC 10646" (see **Annex F**).

2.2.5. Quality Assurance: The 5% Rule

For any character encoding standard, a common general principle is to encode the same character once and only once. Before any submission, it is the submitter's responsibility to filter out the ideographs that are already in the ISO/IEC 10646 international coding standard:
- the published standard,
- any of its published amendments,
- any of its amendments under ballot in JTC1/SC2, or
- any of the working sets of the IRG.

In assessing the suitability of a proposed ideograph for encoding, the IRG will evaluate the credibility and quality of the submitter's proposal. If the IRG finds more than 5% of the submitter's source set are duplicates of characters in the above mentioned collections during the IRG review process, the whole submission will be removed from the subsequent IRG working drafts for that particular IRG project. However, If the proposer explicitly raise question about unification/dis-unification for concrete cases in the proposal of characters, the 5% rule should not be applied for such cases.

2.3. Principles on Production of IRG Working Drafts

After the IRG accepts submissions based on principles specified in Section 2.2, the development process of the review work begins. The IRG technical editor will first produce a set of IRG working drafts.

2.3.1. Principles on Submitted Ideographs

a. All the original ideograph submissions, including glyphs, IDS, radicals, stroke counts and evidence, must have registered IRG document numbers.
b. If any required information is missing, the IRG chief editor or technical editor can ask for additional information from the submitter. Without timely supply of such information, the submission may be rejected by the technical editor in producing a working draft.

2.3.2. Principles on Assignment of Serial Numbers

a. The IRG technical editor will consolidate and sort the submitted ideographs in accordance with Annex A of this document.
b. A unique *serial number* will be assigned to each submitted ideograph after consolidation. The serial numbers must be unique throughout the entire standardization process. They must not be changed, re-set or re-assigned unless an there is an agreed disunification during the process. This principle allows easy reference to past discussions. In case of a split, one ideograph will keep the original serial number and a new serial number will be assigned to the split ideograph.
c. If ideographs submitted by different member bodies are obviously unifiable, such ideographs may be unified and assigned the same serial number by the IRG technical editor.

2.3.3. Principles on Machine-checking of IDS of Submitted Ideographs

a. The IRG technical editor will check the submitted IDS with existing IDS data to detect possible unifiable or duplicated ideographs.

Deleted: a split happens

b. Machine checking sometimes detects obviously non-unifiable pairs. Such cases, when detected, will be annotated before proceeding to the next stage.

c. IDS checking algorithm will satisfy the requirements described in Annex B.

2.3.4. Production of IRG Working Drafts

a. **Division of Character Subsets**: By the result of IDS checking, submitted ideographs will be grouped into the following two working sets:

   i. **M-set (main set)**: for ideographs with proper IDS, and found not to be unifiable with current standardized ideographs nor previously discussed ideographs with proper IDS.

   ii. **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 and supported by evidence for dis-unification.

b. **Naming of Working Drafts**: The file name should follow the format of "IRGN*nnnn*V*X*[*XXX*]" where *nnnn* is the IRG assigned document number and *X* is the version number. No space is allowed. But, the use of underscore "_" and period "." for separation is allowed. Examples of version numbers are "ExtFV1.0", "V1.0Draft", etc.

c. **Glyph Images**: Archive of consolidated glyph images whose image size should be 128x128 with file name using the Source reference with the extension .bmp.

d. **Addition of Characters**: No ideographs should be added to the working sets once the development process begins.

e. **Alteration of Characters:** Generally speaking, alteration of characters indicates instability and any change may also have impact on other characters in the collection. Thus it is generally not allowed. However, member bodies may submit minor alteration of characters with provision of justification ONLY at the final stage or with explicit approval from IRG as long as the alteration is unifiable with the original character. Change of glyph beyond the Annex S unification criteria is considered to be an addition of new character and is NOT acceptable during the development process. The submitter must provide the results of thorough checks and verification that the alteration does not affect other characters in existing standards and working sets. The IRG, based on its evaluation, may decide to accept the alteration, reject the alteration or request the removal of such a character by the submitter. If the submitter finds that the glyph of a character is wrong at any working stages, the character will be rejected by the IRG and should be withdrawn by the submitter.

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.

2.4. Principles on Reviewing IRG Working Drafts

If the IRG instructs member bodies to review a working draft, member bodies' editors should review it (different portions may be assigned to different member bodies) according to the agreed schedule and they should follow the principles set out below during the review process.

2.4.1. General Principles on Reviews

a. Each member body should check the ideographs of the working sets assigned by the IRG chief editor and technical editor for the following issues:

   i. Correctness of KangXi radical and KangXi Index, Stroke Count, First Stroke and IDS.

   ii. Correctness and quality of glyphs and source information if necessary.

   iii. Any duplicate or unifiable ideographs based on Annex S guidelines.

   iv. Consistency of submitted characters with the submitted evidence and documentary proof.

b. When any data, including IDS, KangXi radical, or stroke count are found to be incorrect, such M-set ideograph should be moved to D-set as its standing data are no longer valid. Until such ideograph is assured to be unique by manual checking (procedures described in Section 2.4.2. below), it should not be moved back to M-set.

Deleted: **Previous D-Set:** If a previously discussed D-set exists, new D-set ideographs should be merged with the previous D-set. …

Deleted:

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 unified with any ideographs in the standard or in another working set (including the current one).

b. **Radical Checking:** Assurance is done by enumerating all possible radicals of a target ideograph and looking for any duplicate or unifiable ideographs in the range of ±2 stroke counts of standardized and working set ideographs. For example, "聞" may have the radical of "門" with 6 strokes, or the radical of "耳" with 8 strokes.   In such a case, checking standardized and working set ideographs with radical of "門" and 4-8 strokes, or ideographs with radical of "耳" and strokes of 6-10 manually can have much better assurance that such an ideograph does not have duplicate or unifiable ideographs.

c. **Recording of Review Results**: After reviewing, the reviewer should record the comment of "Checked against all standardized and working set ideographs with radical *X* and stroke count of *Y*±2."

2.4.3. Submission of Possibly Unifiable Ideographs

a. **Preparation of Comments**: Member bodies should prepare comments and feedback with reference to the assigned serial number of the ideograph in question. The guidelines on comments are described in Section 4 of this document. Comment files should be in CSV form as a text file or a Microsoft Excel format file. All comment files must have pre-assigned IRG document numbers.

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 further evidence (for example, from dictionaries, legal documents or other publications) showing that it is not unifiable with another standardized ideograph or an ideograph proposed in the same or another working draft.

c. **Submission Deadline:** Each member body should send feedback comments at least two months before the next IRG meeting.   The IRG chief editor and technical editor will consolidate them and register the result as IRG documents a month before the next IRG meeting so that each member body can examine the comments and prepare any additional documents for discussion at the meeting.

d. **Rejection**: Questioned ideographs with no counter arguments in support of dis-unification supplied to the meeting will be automatically marked as unified.

2.5. Principles on Discussions at IRG Meetings

2.5.1. Document-based Discussion

For efficient and smooth work, all discussion items and evidence must be prepared with registered IRG documents before the commencement of an IRG meeting.   Items or evidence that are not contained in an IRG registered document are not treated as evidence and will not be discussed during IRG meetings. Any discussions on evidence or items raised after the commencement of an IRG meeting may be postponed to the next IRG meeting if any member body requests longer time to examine such items or evidence.

2.5.2. Discussion Procedures

Discussion will be based on the review comments on working sets. For non-unification issues, a submitter should present evidence document(s) showing that suspected unifiable ideographs are distinctively used as non-cognate characters in the same region, or that these two characters cannot be unified in accordance with Annex S.   When IRG member bodies have consensus that the ideographs are unifiable, the submitter should take one of the following actions, and the decision must be recorded.

a. Withdraw the duplicate ideograph and map the character in question to the existing standardized or working set ideograph.

b. Submit it as a compatibility ideograph character.

c. Add a new source reference to the existing standardized or working set ideograph.

When characters are reviewed by different people, different choices of radical, stroke count or first stroke code are possible for the same ideograph. IRG member bodies should resolve to agree on

the most appropriate one based on the commonest abstract shape of the specific glyph.   When KangXi radical or stroke count is found to be incorrect, the ideographs will be moved to D-set and wait for another manual review to prevent any unification error caused by not having conducted the review with ideographs having the correct KangXi radical or stroke count.

Guidelines on typical comments and resolutions are given in Section 4 of this document.

2.5.3. Recording of Discussions

Comments, rationales, and decisions must be recorded for each ideograph reviewed in a tabular format for reference and checking.

2.5.4. Time and Quality Management

Before discussion begins, the number of ideographs under review will be counted and the estimated schedule will be determined based on it. During the discussion, the number of comments reviewed per hour will be noted and the schedule will be adjusted by this rate (Note: It is recognized that some comments may take longer than others to discuss and resolve). If the comments cannot be handled in one IRG meeting, they may be partitioned and resolved in subsequent IRG meetings. Due to the limited time CJK Editorial Group has to deal with individual characters during an IRG meeting, member bodies can use emails to discuss and reach agreement on simple, straightforward cases before and after an IRG meeting.

2.6. Principles on Submission of Ideographs to WG2

2.6.1. Checking of Stabilized M-Set

a.   Once M-set is consolidated and stabilized, the ideographs in M-set will be checked intensively as a complete set at least once to ensure data and glyph integrity.
b.   Approval by IRG using majority vote by member bodies is needed before the set can be prepared for WG2 submission.

2.6.2. Preparation for WG2 Submission

After the approval by   IRG, the IRG technical editor will prepare the proposal to be forwarded to WG2. The preparation includes the following:
a. Sort the final stable M-set ideographs by the sorting algorithm described in Annex A.
b. Assign provisional UCS code positions to the sorted M-set ideographs (with agreement from ISO/IEC 10646 project editor on block assignment).
c. Make available the TrueType fonts for each member body with assigned provisional UCS code positions (fonts have to be available in accordance with the requirement stated in point 5 of A.1. – Submitter's Responsibilities in Annex A, WG2 Principles and Procedures).
  i.   Each submitter is encouraged to prepare and submit its own font for best font quality.
  ii.   If a submitter has difficulties in creating the font, other member bodies or the IRG technical editor may help create the font. In this case, the glyph style of the submitter must be respected.
  iii.   If the submitter cannot provide the TrueType font by this time, the collection by the submitter will be withdrawn from this working set.
d. Prepare a list of source references.
e. Produce a packed Multi-column Ideograph Chart using the TrueType fonts.

The IRG will conduct at least one round of review of the proposal and the chart generated using TrueType font before submission to WG2.

3. Procedures

This section describes the basic development procedures of CJK Unified Ideograph extensions. The ultimate purpose of the procedures outlined in this section is to realize the production of high quality CJK Unified Ideograph sets in an efficient manner.

Development procedures described in this section consists of 8 stages, and it may take two to three years to create a high quality ideograph set for standardization.

3.1. Call for Submission

a. When a member body requests a new project for CJK Unified Ideograph extension and when the project is agreed upon at an IRG meeting, the IRG may call for submission of new ideographs.   The IRG will also determine the deadline for submission.
b. Each member body with proposed ideographs must submit the ideographs before the specified deadline with required data described in Section 2 of this document.
c. Member bodies must check whether the submitted ideographs are accompanied with all required information. If some required information is missing or misplaced, the IRG chief editor or technical editor may ask the submitter to re-submit or supply the additional information if only minor problems are encountered. Otherwise, the submission may be rejected because consolidation with other member bodies' submissions cannot be carried out.

3.2. Consolidation and Grouping of Submitted Ideographs

Consolidation of submissions is normally done between IRG meetings. The consolidation includes the following tasks:
a. The IRG technical editor will sort and assign *serial numbers* to submitted ideographs as described in Section 2.3.2.
b. After serial numbers are assigned, submitted ideographs must undergo IDS checking to detect any duplication and unification.   By 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.
c. After consolidation, a working draft will be assigned an IRG N document number with a version number, and will be distributed to member bodies' editors and made available on the official website of the IRG so that any other experts can have access to it.   The IRG chief editor and technical editor may assign member body's editors to check M-set and D-set ideographs either for the entire collection or certain portions of it depending on reasonable estimation of workload by the IRG chief editor and technical editor.

3.3. First Checking Stage

This stage, which is between IRG meetings, involves the following tasks:
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.
b. Member bodies must submit their comments to the IRG chief editor and technical editor at least two months before the next IRG meeting or according to the IRG approved working schedule.
c. The IRG chief editor and technical editor must consolidate the comments and produce an IRG registered document for circulation and discussion at least one month before the next IRG meeting or according the IRG approved working schedule.
d. Submitters and outside experts are encouraged to prepare and submit supplementary documents (with IRG document numbers) so that they can be discussed at the next IRG meeting.

3.4. First Discussion and Conclusion Stage

This stage, which is during an IRG meeting, includes the following tasks:
a. Member bodies and participating experts should review the comments which are officially submitted before the meeting with assigned IRG document numbers and the editorial group must reach conclusion for each commented ideograph in writing. Guidelines for typical conclusions are provided in Section 4.
b. All the conclusions must be agreed to 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.
c. The IRG technical editor will create a new M-set and D-set a month after the IRG meeting, and register them as IRG registered documents with version information.

    d.   If more than 5% of ideographs submitted by a specific submitter are removed as a result of duplication or unification with existing standardized sets, the entire submission of this submitter will be removed to ensure high quality of the project. This is known as the 5% Rule described in Section 2.2.5 above.

    e.   If new unification or non-unification cases or rules are agreed, IRG should instruct the IWDS editor to modify IRGWDS to reflect the agreement for future works." "5% rules will not be considered for the unification based on newly agreed unification rule.

3.5. Subsequent Checking Stage

Deleted: Second

This stage, which is between IRG meetings, involves the following tasks:
    a.   Each member body's editor must check the newly created M-set and D-set for correctness and duplication.
    b.   Member bodies should submit their comments with registered IRG document numbers to the IRG chief editor and technical editor at least two months before the next IRG meeting or according to the IRG approved working schedule.
    c.   The IRG chief editor and technical editor will consolidate the comments and produce a registered IRG document for circulation and discussion at least a month before the next IRG meeting or according to the IRG approved working schedule.
    d.   Member bodies and outside experts are encouraged to prepare and submit supplementary documents to facilitate discussion during the next IRG meeting.

3.6. Subsequent Consolidation and Conclusion Stage

Deleted: Second

This stage, which is during an IRG meeting, includes the following tasks:
    a.   Member bodies must review the comments and draw conclusion for each ideograph. Typical comment and conclusion examples for each set are provided in Section 4.
    b.   All the conclusions must be agreed to 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.
    c.   The IRG technical editor will create a new M-set and D-set a month after the IRG meeting, and produce an IRG registered document.
    d.   If more than 5% of the ideographs submitted by a specific submitter are removed as a result of duplication or unification with existing standardized sets, the entire submission of this submitter will be removed to ensure high quality of the project. This rule would not apply to newly added unification rules added after the first checking stage.

3.7. Final Checking Stage

This stage, which is between IRG meetings, involves the following tasks:
    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.
    b.   Member bodies' editors must submit their comments to the IRG chief editor and technical editor at least two months before the next IRG meeting.
    c.   The IRG chief editor and technical editor will consolidate the comments and produce an IRG registered document for circulation and discussion at least a month before the next IRG meeting so that member bodies' editors can have time to review them before the next IRG meeting.

3.8. Approval and Submission to WG2

This stage, which is during an IRG meeting, involves the following tasks:
    a.   Member bodies should review the comments on M-set and reach conclusion for each ideograph.
    b.   If there is no positive decision on an M-set ideograph, it will be moved to D-set. 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.

IRG N1920Draft3   **IRG Principles and Procedures**
2013-05-22   Version 6 Draft3   **Page 14 of** 36

c. With the approval from the majority of IRG member bodies, M-set will be frozen as the new ideograph extension set to be submitted to WG2. The IRG technical editor will prepare the document in accordance with Section 2.6 of this document.

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

4. Guidelines for Comments and Resolutions on Working Sets

The following tables list guidelines for typical comments and conclusions during the development process. All comments must be accompanied with date (in YY-MM-DD format) and member body abbreviation(G, T, H, M, J, K, KP, MY, U, V or Z). All conclusions must also be dated.

4.1. Guidelines for M-Set

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 working sets altogether.

When a comment claims that component(s) of a glyph is wrong, the relevant component(s) of ideographs need to be indicated in red circles/boxes in comment files as shown below.

Similarly, when a comment claims that components of two or more ideographs are the same or different, the corresponding components of two or more ideographs need to be indicated in red circles/boxes in comment files as shown below.

The table below also shows samples of possible comments by reviewers and the possible actions associated with these problems

Possible Comment by a Reviewer	Possible Resolution
Wrong or Missing Glyph	• Glyph is corrected, or the missing glyph is supplied. The ideograph is moved to D-set for manual checking.
Wrong KangXi radical / strokes count / first stroke	• Data will be corrected and this Ideograph will be moved to D-set for further manual checking.
Wrong IDS	• IDS will be corrected and the character will be moved to D-set until it is checked again by the IDS checker. • Move to D-set (in case IDS cannot be corrected).
May be unifiable with U+xxxxx (standardized ideograph)	• Unified with U+xxxxx and the submitter will request a new source reference to U+xxxxx. • Unified with U+xxxxx and the submitter will request that this character be treated as a Compatibility Ideograph.

	● Unified to U+xxxxx and this entry will be removed. (May consider to register it to IVS.) ● Not unifiable.
May be unifiable with xxxxx (M-set ideograph)	● Unified with xxxxx and this source reference will be attached to xxxxx. ● Unified with xxxxx and the submitter may consider registering it as a Compatibility Ideograph Character or IVS. ● Not unifiable.

4.2. Guidelines for D-Set

Ideographs in D-Set are either the ones that cannot be checked automatically by IDS checking algorithm or the ones of which the attribute data have been questioned by a member body or that are suspected to be unifiable with other standardized or working set ideographs. For the ideographs that cannot be machine-checked by IDS matching, at least two non-submitter member bodies must check them manually to ensure that the ideographs are not unifiable with any standardized ideographs or working set ideographs. For the ideographs that might be unifiable with other ideographs, the submitters of these ideographs are requested to prepare arguments and evidence to show that such ideographs should be separately encoded.

Possible Comment by IDS Checker	Possible Conclusion
● Incomplete IDS IDS with extra character ● Component is not an ideograph	● IDS will be corrected and it will be moved to M-set when next IDS-check is done. ● Proper IDS cannot be generated and manual checking is needed.
Possible Comment by a Reviewer	**Possible Conclusion**
● Wrong KangXi radical ● Wrong stroke count ● Wrong first stroke	● Data will be corrected. ● Proposal to correct data is not accepted, as it is an ambiguous case and the IRG agrees that the previous choice of *XX* is more appropriate.
● Wrong IDS	● IDS will be corrected and will be checked by the IDS checker again. ● Correct IDS cannot be generated and manual checking is needed.
May be unifiable with U+xxxxx (standardized ideograph)	● Unified with U+xxxxx and a new source will be added to U+xxxxx. The new candidate entry should be deleted. ● Not unifiable, as shown by the evidence *IRG N xxxx*. Move to M-set.
May be unifiable with *xxxxx* (M-set or D-set ideograph)	● Unified with *xxxxx* in M-set and a new source will be added to *xxxxx*. The new candidate entry should be deleted from D-Set. ● Unified with *xxxxx* in D-Set and a new source will be added to *xxxxx*. The new candidate entry should be removed from D-Set. ● Not unifiable, as shown by the evidence *IRG N xxxx*. Move to M-set
Checked against all standardized and working set ideographs with radical *X* and stroke count of *Y*±2.	● Move to M-set, as two non-submitter member bodies (*XX* and *YY*) confirmed that this ideograph is not unifiable with any existing standardized or working set ideographs. ● Checking against ideographs with radical *X* may not be enough. This ideograph will also be checked against ideographs with radical *Z*.

4.3. Guidelines for comments

a. When a comment claims that component(s) of a glyph is wrong, the relevant component(s) of ideographs need to be indicated in red circles/boxes in comment files.

Similarly, when a comment claims that components of two or more ideographs are the same or different, the corresponding components of two or more ideographs need to be indicated in red circles/boxes in comment files.

5. IRG Website

The IRG maintains its own web site at http://www.cse.cuhk.edu.hk/~irg/, hosted by the Department of Computer Science and Engineering at The Chinese University of Hong Kong. IRG meeting notices, resolutions, document register, documents and standing documents are made available at this site. Hyperlinks to WG2 websites will be provided for member bodies' easy access. For faster retrieval of documents and searching, documents should not be compressed as far as possible and the site search engine window should be made available. Documents larger than 4MB must be split into multiple files for easy uploading, downloading and searching. The compressed files must be in WinZip format with .zip extension.

6. IRG Document Registration

All documents to be formally discussed by the IRG must be registered with assigned IRG document numbers (assigned by the IRG Rapporteur) and contain submission date, title, submitting member body, or the author, purpose (or summary), and the 'IRG Ideographic Repertoire Submission Summary Form' (when applicable).

6.1. Registration Procedures

The following gives the registration procedures:
   a. **Request for Document Number:** All documents submitted to the IRG must be given a registered document number. The assignment is done by the IRG Rapporteur. A member body will first contact the IRG Rapporteur for a document number with a document title. Once the document number is assigned, the information will be posted on the IRG website. Some document numbers can be pre-assigned during IRG meetings for activities between IRG meetings.
   b. **Submission of Documents:** All registered documents must be submitted to the IRG Rapporteur. The submitted documents must also contain an assigned IRG document number in text form(except files of pure tables to avoid interfering with the data presented in the table) so that searching can be supported.
   c. **Posting of Documents**: Properly submitted documents are then posted by the IRG Rapporteur on the IRG website as official documents.
   d. **Disqualified Documents**: Documents with certain basic information missing such as submitter's name, title and purpose may be rejected by the IRG Rapporteur for posting. All other documents which fail to comply with the above registration process and the preliminary review by the IRG Rapporteur for basic information will not be treated as IRG documents. As such, issues contained in such documents will not be discussed by the IRG formally.

6.2. Contact for IRG Document Registration

The current IRG Rapporteur is Prof. Qin LU and her contact information is as follows:

Professor Qin Lu
Department of Computing
The Hong Kong Polytechnic University
Hung Hom, Hong Kong
Tel. (852) 2766 7247
Fax. (852) 2774 0842
Email: csluqin@comp.polyu.edu.hk

Annex A: Sorting Algorithm of Ideographs

The IRG recognizes that the choice of radicals, the sequence of strokes, and the choice of strokes are locale dependent. Thus, submitters may have different preferences of character orders. However, for the convenience of IRG editorial work, the IRG must adopt a sorting order for the convenience of editorial which which may be different from submitters' preference. Thus the principles of sorting of the ideographs given below are internal for IRG editing purposes only. Ideographs consolidate for unification review must be sorted according to the following order.

a.   **KangXi Radical Order**.
   Note: When radicals are in simplified forms given below, ideographs with simplified radicals must be placed after the ideographs with corresponding traditional radicals.

Traditional Radicals		Simplified Radicals	
R119.0	糸	R119.1	纟
R146.0	見	R146.1	见
R148.0	言	R148.1	讠
R153.0	貝	R153.1	贝
R158.0	車	R158.1	车
R166.0	金	R166.1	钅
R167.0	長	R167.1	长
R168.0	門	R168.1	门
R177.0	韋	R177.1	韦
R180.0	頁	R180.1	页
R181.0	風	R181.1	风
R182.0	飛	R182.1	飞
R183.0	食	R183.1	饣
R186.0	馬	R186.1	马
R194.0	魚	R194.1	鱼
R195.0	鳥	R195.1	鸟
R196.0	鹵	R196.1	卤
R198.0	麥	R198.1	麦
R204.0	黽	R204.1	黾
R209.0	齊	R209.1	齐
R210.0	齒	R210.1	齿
R211.0	龍	R211.1	龙

b.   **Stroke Count**.
   Note: Simplified characters must be placed after traditional characters within the same stroke-number group.

c.   **First Stroke**
   The technical editor will assign the first stroke based on IRGN954AR and IRGN1105. In case of previous unseen components, the technical editor will take the conventions of Kangxi for first stroke assignment without regards to the submitter's locale conventions.

Annex B: IDS Matching

B.1. Guidelines on Creation of IDS

Each member body should consult IRG N1183 on IDS. It should be noted that in addition to the CDC (Character Description Components) defined in IRG N1183, all CJK Unified Ideographs accepted by ISO/IEC 10646 in its amendments are also qualified as CDC in constructing IDS.

The use of "overlapping" IDC(Ideographic Description Characters) or more than four IDCs is considered to be 'inappropriate' and may not be a subject of IDS comparison.

B.2. Requirements on IDS Matching

The IDS matching algorithm used by the IRG should support the following features:

1. IDS matching should be able to handle different split points.
   (e.g. ⿰⿰亻頃 and ⿰化頁 should be matched.)
2. IDS matching should be able to handle different split levels.
   (e.g. ⿰⿰亻悉 and ⿰⿰亻⿱釆心 should be matched.)
3. IDS matching should match different glyphs of the same abstract shape.
   (e.g. ⿰⿰礻申 and ⿰示申 should be matched.)
4. IDS matching should match similar glyphs.
   (e.g. ⿰⿰忄生 and ⿰小生 should be matched.)
5. IDS matching should match IDS with different orderings of overlapping IDC.
   (e.g. ⿷三丨 and ⿷丨三 should be matched.)
6. IDS matching should match unifiable IDC patterns.
   (e.g. ⿰⿰麥离 and ⿸麥离 should be matched.)
7. IDS matching should be able to handle any combination of the above.
8. IDS matching should be able to detect any inappropriate IDS, such as IDS being too long, IDS with non-ideographic CDC, or missing or extra CDC or IDC.

B.3. Limitation of IDS Matching

It should be noted that IDS matching cannot detect unification or duplication if a component cannot be encoded by an IDS, or if the glyph itself is very complex. IDS matching is done algorithmically. It is not versatile on the detection of unifiable ideographs unless rules are explicitly given to the algorithm. Thus, it is not meant to be the replacement of manual checking. Rather, it is an assistive tool for quality assurance to identify duplication and known cases of unification. Therefore, it is very important for submitters to make sure that their submitted ideographs are not going to be unified with any standardized or previously discussed ideographs or working set ideographs.

Deleted: creation finalized at IRG Meeting No. 25

Annex C: Urgently Needed Ideographs

C.1. Introduction

When a member body demonstrates an urgent need for a small number of ideographs to be standardized rapidly for justifiable reasons, such as ideographs in a recently developed regional or national standard that must be implemented by a particular deadline, the IRG may submit the ideographs, independent of any of the current IRG working sets to WG2. Each member body's urgently-needed submission shall be treated as a separate urgently-needed repertoire, and a member body can have no more than one active urgently-needed submission and the process will be started only sparingly with demonstrated real need.

C.2. Requirements

A member body that is submitting urgently-needed ideographs must prepare the following:

a. All the documents required for normal ideograph submissions.
b. In addition to the above, a document that indicates whether, among the submitted urgently-needed ideographs, there are any that can be unified with ideographs in the current IRG working sets. When a particular urgently-needed ideograph repertoire is accepted by WG2, any unifiable ideographs in the current working sets will be removed as explained in C.3 below.
c. For ideographs not mentioned above, the document must prove that their submitted urgently-needed ideographs are not unifiable with any ideographs in the current working sets. The proof may be provided by listing the documents the submitter has checked, and for each proposed ideograph, a list of ideographs whose radicals and strokes were checked against. It is an important responsibility of the submitter to check with not only the current standardized CJK ideographs, but also the IRG working sets for any unifiable characters against its submission. If a submitter fails to do the above, the submission will not be approved by the IRG as an IRG-endorsed independent submission to WG2.
d. A submission is deemed urgently-needed only if the submitting member body demonstrates urgency or a rationale for rapid standardization.
e. Each submission from a member body shall include no more than 30 ideographs. Submissions that include more than 30 characters will be accepted at the sole discretion of the IRG.

C.3. Dealing with Urgent Requests

Accepted urgently-needed ideographs as independent submissions must be checked by the IRG for correctness, duplication and unification for the latest published ISO/IEC 10646 as well as the current IRG working sets. When an ideograph is found to be identical or unifiable with the ones in the current IRG working sets, such ideograph must be noted and removed from the current IRG working sets.

Annex D: Up-to-date CJK Unified Ideograph Sources and Source References

The IRG tracks the sources of the CJK Unified Ideographs. The past practice is that the characters sources are tracked based on submissions by member bodies. Thus each member body has a member body abbreviation as defined by ISO/IEC 10646. In recognizing that submitters may be an international consortium not affiliated with any member body, the IRG has decided in its IRG Meeting No. 39 to add a new abbreviation Z used for this type of submission. It is noted that Z source may include submissions from different projects and additional letters may be used to indicate each repertoire.

D.1. Member body abbreviations:
G	China
H	Hong Kong Special Administrative Region, China
J	Japan
K	Republic of Korea
KP	Democratic People's Republic of Korea
M	Macao Special Administrative Region, China
MY	Malaysia (Added in Nov. 2008 at IRG Meeting No. 31).
T	Taipei Computer Association
U	Unicode Consortium
V	Vietnam

Note: all member body abbreviations except MY are currently used in ISO/IEC 10646 Section 23. MY is defined and used by IRG internally only.

D.2. The Hanzi G sources
G0	GB2312-80
G1	GB12345-90 with 58 Hong Kong and 92 Korean "Idu" characters
G3	GB7589-87 unsimplified forms
G5	GB7590-87 unsimplified forms
G7	General Purpose Hanzi List for Modern Chinese Language, and General List of Simplified Hanzi
GS	Singapore Characters
G8	GB8565-88
G9	GB18030-2000
GE	GB16500-95
GH	GB15564-1995 Code of Chinese Ideogram set for teltext broadcasting Hong Kong subset
GK	GB12052-89 Korean Character Coded Character Set for Information Interchange
G_4K	Siku Quanshu （四庫全書）
G_BK	Chinese Encyclopedia （中國大百科全書）
G_CH	Ci Hai （辞海）
G_CY	Ci Yuan （辭源）
G_CYY	Chinese Academy of Surveying and Mapping Ideographs (中国测绘科学院用字）
G_FZ	Founder Press System （方正排版系统）
G_GH	Gudai Hanyu Cidian （古代汉语词典）
G_HC	Hanyu Dacidian （漢語大詞典）
G_HZ	Hanyu Dazidian ideographs（漢語大字典）
G_IDC	ID system of the Ministry of Public Security of China, 2009
G_GJZ	Commercial Press Ideographs （商务印书馆用字）
G_K	GKX Kangxi Dictionary ideographs（康熙字典）9th edition (1958) including the addendum （康熙字典）補遺
G_GXC	Xiandai Hanyu Cidian (现代汉语词典）
G_ZFY	Hanyu Fangyan Dacidian （汉语方言大辞典）
G_ZH	Zhonghua Zihai (中华字海)
G_ZJW	Yinzhou Jinwen Jicheng Yinde （殷周金文集成引得）
G_XH	Xinhua Zidian (新华字典)
G_WZ	Hanyu Dacidian Publishing House Ideographs（漢語大詞典出版社用字）

D.3. Hanzi H sources
H	Hong Kong Supplementary Character Set (HKSCS)

D.4. Hanzi T sources
    T1   TCA-CNS 11643-1992 1st plane
    T2   TCA-CNS 11643-1992 2nd plane
    T3   TCA-CNS 11643-1992 3rd plane with some additional characters
    T4   TCA-CNS 11643-1992 4th plane
    T5   TCA-CNS 11643-1992 5th plane
    T6   TCA-CNS 11643-1992 6th plane
    T7   TCA-CNS 11643-1992 7th plane
    TB   TCA-CNS 11643-2007 11th plane
    TC   TCA-CNS 11643-2007 12th plane
    TD   TCA-CNS 11643-2007 13th plane
    TE   TCA-CNS 11643-2007 14th plane
    TF   TCA-CNS 11643-2007 15th plane

D.5. Kanji J sources
    J0   JIS X 0208-1990
    J1   JIS X 0212-1990
    J3   JIS X 0213:2000 level-3
    J3A   JIS X 0213:2004 level-3
    J4   JIS X 0213:2000 level-4
    JA   Unified Japanese IT Vendors Contemporary Ideographs, 1993
    JH   Hanyo-Denshi Program (汎用電子情報交換環境整備プログラム), 2002-2009
    JK   Japanese KOKUJI Collection
    JARIB   Association of Radio Industries and Businesses (ARIB) ARIB STD-B24 Version 5.1, March 14 2007

D.6. Hanja K sources
    K0   KS C 5601-1987
    K1   KS C 5657-1991
    K2   KS X 1027-1:2011 (previously known as PKS C 5700-1 1994)
    K3   KS X 1027-2:2011 (previously known as PKS C 5700-2 1994)
    K4   KS X 1027-3:2011 (previously known as PKS 5700-3:1998)
    K5   KS X 1027-4:2011 (previously known as Korean IRG Hanja Character Set
            5th Edition: 2001)

D.7. Hanja KP sources
    KP0   KPS 9566-97
    KP1   KPS 10721-2000

D.8. ChuNom V sources
    V0   TCVN 5773:1993
    V1   TCVN 6056:1995
    V2   VHN 01:1998
    V3   VHN 02: 1998
    V4   Dictionary on Nom 2006, Dictionary on Nom of Tay ethnic 2006, Lookup Table for Nom in the
            South 1994

D.9. MY sources
    MY   " Dictionary Of Chinese Rustic Language In South-East Asia", written by Xu Yunqiao, published by
            Singapore Shjie Publisher, 1961. 《南洋华语俚俗辞典》, 新加坡世界书局有限公司, 1961 年 8 月

D.10. Macao sources
    MAC   Macao Supplementary Character Set

D.11. Unicode sources
    UTC   The Unicode Technical Report #45, U-source Ideographs, May 2010

D.12. Z sources
==Sat to be added in IRG Meeting No. 40.==

Annex E: Maintenance Procedure of the IRG Working Documents Series

E.1 Introduction

The IRG Working Documents Series is a set of IRG maintained documents which keeps the up-to-date examples of CJK unification related example cases to supplement the published Annex S of ISO/IEC 10646 for IRG unification work.

E.2. IRG Working Documents Series

The document formats and the specific lists are maintained as a separate set of documents as the IRG Working Document Series (IWDS).
Series 1:   Summary of unification rules and sample examples (File name: IWDS_SUM.pdf)
Series 2:   List of UCV (Unifiable Component Variations) of Ideographs (File name: IWDS_UCV.pdf)
Series 3:   List of Non-unifiable Components of Ideographs and Overly-unified Ideographs (File name: IWDS_NUC.pdf)
Series 4:   List of Possibly Mis-unified Ideographs (File name: IWDS_MUI.pdf).

E.3. Maintenance Procedure
The maintenance procedure describes how entries in the IWDS are added, removed, or changed.   The IRG has an appointed IWDS Editor (currently, Mr. Taichi Kawabata) who is in charge of the maintenance of the IWDS.

In principle, all update requests are results of IRG unification review work. A review cycle between two IRG meetings is needed. Every update must be discussed in at least one IRG meeting and confirmed in writing. The update is normally started from the assigned unification review work given to member bodies in the past IRG meeting (Meeting No. N-1). Then, during the review work before the next IRG meeting (Meeting No. N), if member bodies found duplicates, unifiable cases or mistakes, which may warrant a change in the IWDS, they need to report these cases in a specified form attached to IWD Series 1.These reported cases will then be consolidated by the IRG technical editor before IRG Meeting No. N. During IRG meeting No. N, time must be allocated to discuss these reported cases and conclusions must be recorded during this IRG meeting. Based on the confirmed conclusion on IWDS updates, the IWDS editor will update the IWDS documents. Any unclear conclusions will be further discussed in future meetings.

The IRG recognizes that the update of the IWDS may take time, and sometimes it is difficult to find appropriate examples. The IRG, however, requests the IWDS editor to keep a log on the actions carried out based on IRG instructions so that better tracking of changes can be carried out.

Below is the description of the maintenance procedure as a flow chart.
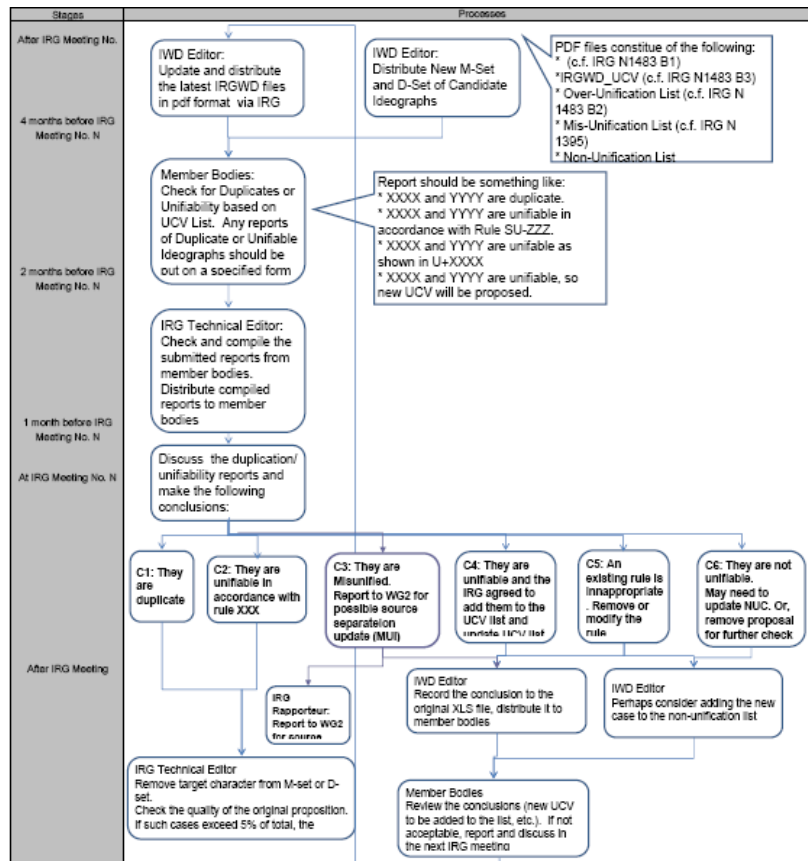
Please refer to the separate Excel file for a more clear diagram.

Annex F: IRG Repertoire Submission Summary Form

ISO/IEC JTC 1/SC 2/WG 2/IRG
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS OF CJK UNIFIED IDEOGRAPHS TO THE REPERTOIRE OF ISO/IEC 10646
Please fill in all the sections below.
Please read Principles and Procedures Document (P & P) from http://appsrv.cse.cuhk.edu.hk/~irg/irg39/IRGN1920.pdf
for guidelines and details before filling in this form.
Please ensure you are using the latest Form from http://appsrv.cse.cuhk.edu.hk/~irg/ irg39/IRGN1920SubmissionForm.xlsx .
See also http://appsrv.cse.cuhk.edu.hk/~irg/irgwds.html **for latest** *Unifiable Calligraphic Variations*.

A. Administrative

1. **IRG Project Code:**	e.g. Extension-F
2. **Title:**	
3. Requester's region/country name:	
4. Requester type (National Body/Individual contribution):	
5. Submission date:	
6. Requested Ideograph Type (Unified or Compatibility Ideographs)	
If Compatibility, does the requester have the intention to register them as IVS (See UTS #37) with the IRG's approval? (Registration fee will not be charged if authorized by the IRG.)	
7. Request Type (Normal Request or Urgently Needed)	
8. Choose one of the following:	
This is a complete proposal:	
(or) More information will be provided later:	

B. Technical – General

1. Number of ideographs in the proposal:	
2. Glyph format of the proposed ideographs: (128x128 "bmp" files or TrueType font file)	
If 'bmp' files, are their file names the same as their source references?	
If TrueType font, are all proposed glyphs put into BMP PUA area?	
If TrueType font, are data for source references vs. character codes provided?	
3. Source references:	
Do all the proposed ideographs have a unique, proper source reference (member body abbreviation followed by no more than 9 alphanumeric characters)?	
4. Evidence:	
a. Do all the proposed ideographs have a separate evidence document which contains at least one scanned image of printed materials (preferably dictionaries)?	
b. Do all the printed materials used for evidence provide enough information to track them by a third party (ISBN numbers, etc.)?	
5. Attribute Data Format: (Excel file or CSV)	

C. Technical - Checklist

Understandings of the Unification Checklist	
1. Has the requester read ISO/IEC 10646 Annex S and does the requester understand the unification policy?	
2. Has the requester read the "Unifiable Calligraphic Variations" (contact IRG technical editor through the IRG Rapporteur for the latest one) and does the requester understand the unifiable variation examples?	
3. Has the requester read this P&P document and does the requester understand the 5% Rule?	
Character-Glyph Duplication Checklist(http://www.itscj.ipsj.or.jp/sc2/open/pow.htm contains all the published ones and those under ballot**)**	
4. Has the requester checked that the proposed ideographs are **not unifiable** with any of the unified or compatibility ideographs of the latest version of ISO/IEC 10646? If the checking has been done against an earlier version of ISO/IEC 10646, please specify the version? (e.g. 10646:2011)	
5. Has the requester checked that the proposed ideographs are **not unifiable** with any of the ideographs in the amendments, if any, of the latest version of ISO/IEC 10646? If yes, which amendments has the requester checked?	
6. Has the requester checked that the proposed ideographs are **not unifiable** with any of the ideographs in the proposed amendments, if any, of ISO/IEC 10646? If yes, which draft amendments has the requester checked?	
7. Has the requester checked that the proposed ideographs are **not unifiable** with any of the ideographs in the current working M-set and D-set of the IRG? (Contact IRG chief editor and technical editor through the IRG Rapporteur for the newest list) If yes, which documents has the requester checked?	
8. Has the requester checked that the proposed ideographs are **not unifiable** with any of the over-unified or mis-unified ideographs in ISO/IEC 10646? (Check Annex E of this document).	
9. Has the requester checked whether the proposed ideographs have any **similar ideographs** in the current standardized or working sets mentioned above?	
10. Has the requester checked whether the proposed ideographs have any **variant ideographs** in the current standardized or working sets mentioned above?	
Attribute Data Checklist	
11. Do all the proposed ideographs have attribute data such as the KangXi radical code and stroke count?	
12. Are there any simplified ideographs (ideographs that are based on the policy described in 簡化 字總表) among the proposed ideographs? If YES, does the proposal include proper simplified/traditional indication flag for each proposed ideograph in the attribute data?	
13. Do all the proposed ideographs have the document page number of evidence documents in the attribute data?	
14. Do all the proposed ideographs have the proper Ideographic Description Sequence (IDS) in the attribute data? If NO, how many proposed ideographs do not have the IDS?	
15. If the answer to question 9 or 10 is yes, do the attribute data include any information on similar/variant ideographs for the proposed ideographs?	

Annex G: Examples of new CJK Unified Ideographs Submissions (i.e., Vertical Extension)

G.1. Sample Data Files

All submitted characters must follow the submission format given in Section 2.2.1. The following gives a list of samples submitted by ROK for consideration in CJK Ext. F work (IRGN 1887).

Seq. No.	a) Source reference	b) Glyph Image file name	c) KangXi Radical code	d) Stroke Count	e)First stroke	f) T/S (traditional 0, or simplified 1)	g) IDS (Ideographic Description Sequence)	h)Similar/ Variants	i)Ref. to Evidence Doc	i1) file name of evidence image
1	KA-KC00001	KA-KC00001.bmp	780	5	1	0	𭕊一𮯪中丶	No	韓國學中央研究員 宮中文化	KA-KC00001.jpg
2	KA-KC05982	KA-KC05982.bmp	780	5	3	0	𭕊千止 / 𭕊丿上 止	No	韓國文集叢刊 259集 青莊館全書	KA-KC05982.jpg
3	KA-KC06969	KA-KC06969.bmp	830	8	5	0	一𭕊乃丶囱	No	韓國文集叢刊 261集 貞蕤閣集	KA-KC06969.jpg

G.2. Sample Evidence

All character submissions must include evidence of use as specified in Section 2.2.3. The following shows an example of a Japanese submission with reference to the use of the character in ancient books (IRG N1225 Part2).

G.3. Handling of Data with Privacy Concerns
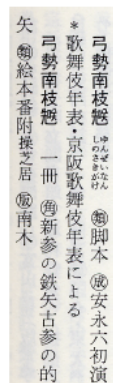
The IRG understands that the current privacy laws and practices in different Countries and Regions can make the submission of complete records as evidence related to personal information may be difficult. As a compromise, the IRG suggests member bodies to provide evidence in such a way that it would not reveal complete personal/internal information. However, the character information itself must be shown in the supplied evidence. In other words, partial document images should be supplied with certain sensitive information blocked.

As different departments/organizations may have different types of documents, the IRG suggests that, for each type of document, a submitter provides a sample document with private information deleted. A good example is the original Basic Certificate of Family Relation Register in Korea as seen in Fig. G1. The evidence can be submitted as partial data in the form shown in Fig. G2.

Figure G1. The original Basic Certificate of Family Relation Register

Figure G2. A modified Basic Certificate of Family Relation Register (private information such as full name and date of birth has been deleted).

G.4. Consideration for acceptance for characters that cannot be provided in Printed forms

The consideration for acceptance for characters that cannot be provided with evidence in printed forms is not meant to relax the IRG requirement on the provision of evidence for modern use. Rather, it is meant to facilitate e-government initiatives for computerized processing of information. It is under this presumption the IRG is willing to consider acceptance of characters already supported in computer systems that are maintained by a designated government body with wide use by government bodies and citizens for administrative public service.

The IRG recognizes that due to e-government issues, some of the characters included in such systems cannot supply supporting evidence for actual use according to 2.2.3a in this document, yet it is technically and administratively not practical to remove them from the computer systems. Thus IRG is willing to consider their acceptance without actual evidence for already implemented working system only. However, IRG requires the submitting member bodies to provide information on the quality assurance process for the maintenance of the respective character collections. Submitters must supply information on the accessibility of the respective character collection and the respective system, the stability and the traceability of the collection, and the kind of evidence/information needed for approval of character removal, modification and addition by the administrative body of the collection.

[Annex H] Not Used at the Moment

Annex H is purposely left out for the time being so that IRG Annex numbers can be the same as WG2 Annex numbers where the subjects are the same.

Annex I: Guideline for Handling of CJK Ideograph Unification or Dis-unification Errors

(Same as WG2 Principles and Procedures Annex I)
Source: www.dkuug.dk/jtc1/sc2/wg2/docs/principles.html

There are two kinds of errors related to coded CJK unified ideographs.
> Case 1: *to be unified* error - Ideographs that should have been unified are assigned separate code points.
> Case 2: *to be dis-unified* error - Ideographs that should not have been unified are unified and assigned a single code point.   An example of this is the request from TCA in document N2271.

When such errors are found, the following guidelines will be used by WG2 to deal with them.

I.1 Guideline for "To Be Unified" Errors

A. The "*to be unified*" pair will be left dis-unified. Once a character is assigned a code position in the standard, it will not be removed from the standard.
B. If necessary, an additional note may be added to an appropriate section in the standard.

I.2 Guideline for "To Be Dis-unified" Errors

A. The ideographs to be dis-unified should be dis-unified and should be given separate code positions as soon as possible.   These ideographs will have two separate glyphs and two separate code positions.   One of these ideographs will stay at its current encoded position.   The other one will have a new glyph and a new code position[4].
B. For the ideographs that are encoded in the BMP, the code charts in ISO/IEC 10646 are presented in multiple columns, with possibly differing glyph shapes in each column.   The question of which glyph will be used for the currently encoded ideograph will be resolved as follows.   In the interest of synchronization between ISO/IEC 10646 and the Unicode standard, the ideograph with the glyph shape that is similar to the glyph that is published in the "Unicode Charts" will continue to be associated with its current code position.   For the ideographs outside the BMP, the glyph shape in ISO/IEC 10646 and the Unicode Charts are identical and will be used with its current code position.
C. The dis-unified ideograph will have a glyph that is different from the one that retains the current code position.
D. The net result will be an addition of new ideograph character and a correction and an additional entry to the source reference table.

I.3 Discouragement of New Dis-unification Requests

There is a possibility of "cognate dis-unification" requests. They are very similar to new source code separation requests.   This kind of requests will not be accepted disregarding the reasoning behind. The key difference between "TO BE DIS-UNIFIED" and "WILL NOT BE DIS-UNIFIED is as follows.
> a. If a character pair is non-cognate (i.e. their meanings are different), that pair of characters is TO BE DIS-UNIFIED.
> b. If a character pair is cognate (i.e. their meanings are the same but have different shapes), that pair of characters WILL NOT BE DIS-UNIFIED.

Dis-unification requests with the reason of mis-application (over-application usually) of the unification rule should NOT be accepted due to the principle in WG2 resolution M41.11.

Deleted: (dis-unification in some sense, and character name change in some sense also)…

---

[4] Disunification not only requires additional code positions. In some sense, there is also the change of the character names.

Annex J: Guideline for Correction of CJK Ideograph Mapping Table Errors

(Same as WG2 P&P Annex J)
Source: ISO/IEC JTC 1/SC 2/WG 2 N2577 – 2003-09-02

In principle, the mapping table or reference to code points of an existing national/regional standard (in the source reference tables) must not be changed.   But once a fatal error is found, it should be corrected as early as possible, under the following guidelines:

J.1 Priority of Error Correction Procedure
   A. Consider adding a new code position and source-reference mapping for the character in question rather than changing the mapping table.
   B. If the change of mapping table is unavoidable, correction should be done as soon as possible.

J.2 Announcement of Additions to or Correction of Mapping Table
   Once any addition to or correction of the mapping table is made, an announcement of the change should be made immediately.   Usually this will be in the form of a resolution of a WG2 meeting, followed by a subsequent process resulting in an appropriate amendment to the standard.

J.3 Collection and Maintenance of Mapping Tables that are not Owned by WG2
   There are many mapping tables, which are included in national/regional standards or developed by third parties.   These are out of WG2's scope.   Any organization (such as Unicode Consortium) that collects mapping information, maintains it consistently and makes this information widely available is invited and encouraged to do so.

Annex K: List of First Strokes

Below gives the list of first strokes including the glyps, names in English, names in Chinese, and the Pinyins of the Chinese names.

Glyph	Stroke No.	Name	Name in Chinese	Pinyin
一	1	horizontal bar	橫	heng2
丨	2	vertical bar	豎	shu4
丿	3	Slash	撇	pie3
丶	4	Dot	點	dian3
乙	5	Turn	折	zhe2

Annex L: Guidelines to forming the current working set with an upper limit

References

Document numbers in the first column in the following table refer to IRG working documents (ISO/IEC JTC 1/SC 2/WG 2/IRGNxxxx), except where noted otherwise. For documents with no link, you may try http://www.cse.cuhk.edu.hk/~irg/ ; some older documents may only be available in paper form (contact the IRG Rapporteur Prof. Qin LU).

Doc. No.	Title	Source	Date
WG2 N3201	Principles and Procedures for Allocation of New Characters and Scripts and Handling of Defect Reports on Character Names	WG2	2007-03-14
N681	Annex S http://standards.iso.org/ittf/PubliclyAvailableStandards/c039921_ISO _IEC_10646_2003(E).zip	Bruce Peterson and IRG Rapporteur	1999-11-18
N881	CJK Extension C Submission Format	IRG	2001-12-04
N953	Minutes of the Adhoc meeting on submitted documents: N941, N942, N944, N945, N948, N949	CJK ad hoc group	2002-11-22
N954	Report on first stroke/stroke count by ad hoc group	CJK ad hoc group	2002-11-22
N954AR	N954 Appendix: First Stroke / Stroke Count Chart	CJK ad hoc group	2002-11-21
N955	IRG Radical Classification	Ideograph Radical Ad Hoc	2002-11-21
N956	Ideograph Unification	Ideograph Radical Ad Hoc	2002-11-21
N1105	Amendments to IRG N954AR	Macao	2005-01-03
N1183	IDS decomposition principles(Revised by the IRG)	KAWABATA, Taichi	2005-12-28
N1197	Sample evidence for CJK C1 candidates	Japan	2006-05-22
N1372	On Better use of IDS on IRG development process	KAWABATA, Taichi	2007-11-09
SC2 N3933		SC2	

Glossary:

Abstract shape: Ideographic characters are used as symbols to represent different entities and used for different purposes. The same character conceptually can sometimes be written in different actual shapes with minor stroke differences, due to preference, which do not affect the recognition of the character as a unique symbol. These characters having the same abstract shapes are not coded separately because ISO/IEC 10646 is a character (symbol) standard, not a glyph standard. In other words, character glyphs (actual shapes) that are considered to have the same abstract shapes are to be unified under the CJK unification rules (defined in Annex S of ISO/IEC 10646).

As ideographs are formed by both the components and the relative positioning of the components, the examination of glyph difference is observed by taking into consideration the meaning, components, and their relative positions. Characters having different meanings and different actual shapes are not considered to have the same abstract shapes. Characters having the same components yet different in relative positions are generally considered to have different abstract shapes. However, component difference is subjected to examination by experts to see if they have influenced the recognition of the character as a whole with consideration of the character's origin and use. Annex S of ISO/IEC 10646 has defined the examination procedure which is given below:

"*The following features of each ideograph to be compared are examined:*
a) the number of components,
b) the relative position of the components in each complete ideograph,
c) the structure of corresponding components.

If one or more of the features a) to c) above are different between the ideographs in the comparison, the ideographs are considered to have different abstract shapes and are therefore not unified.

If all of the features a) to c) above are the same between the ideographs, the ideographs are considered to have the same abstract shape and are therefore unified."

Please also refer to Annex S in ISO/IEC 10646 for examples of characters and components that are considered to have the same abstract shape. The IRG maintains an up-to-date Unification Examples List.

Character Description Component: It refers to any symbols that can be used with the Ideograph Description Characters to form a Ideograph Description Sequence. It includes all coded CJK unified ideographs, Kangxi Radicals, CJK Radical Supplements, and coded CJK Compatibility ideographs.

CJK Unified Ideographs.   It refers to the collection of unified Han characters in ISO/IEC 10646 standard. CJK stands for Chinese, Japanese and Korean. The term CJK Unified Ideographs was adopted at the earlier years of the IRG to reflect the development work of the Han character unification from the three languages at that time.   It is obvious today that Han unification covers far beyond the scripts used in China, Japan and Korea. However, the name is consistently being used in the standardization process and is not changed.

Compatibility Ideographs: Compatibility ideographs are a kind of compatibility characters defined in Section 18 of ISO/IEC 10646. Below is a direct quote from ISO/IEC 10646-2012:

"The CJK compatibility ideographs are ideographs that should have been unified with one of the CJK unified ideographs, per the unification rule described in Annex S. However, they are included in this International Standard as separate characters, because, based on various national, cultural, or historical reasons for some specific country and region, some national and regional standards assign separate code points for them."

D-set (discussion set): D-set is the set of characters that have been reviewed by IRG member bodies with pending issues which need further discussion/evidence for inclusion in the M-set of a working set.

Ideographic Description Characters (IDC): The 12 characters defined in ISO/IEC 10646 starting from the code point U+2FF0: ⿰⿱⿲⿳⿴⿵⿶⿷⿸⿹⿺⿻.

Ideographic Description Sequence (IDS):   IDS describes a character using its components and indicating the relative positions of the components. IDCs are considered operators to the components.

IDSs can be expressed by a context free grammar through the Backus Naur Form (BNF). The grammar G has four components:

Let G = {Σ, N, P, S}, where
- Σ: the set of terminal symbols including all coded radicals and coded ideographs(referred to as CDC, Character Description Components), and the 12 IDCs.
- N:the set of 5 non-terminal symbols
        N={IDS, IDS1, Binary_Symbol, Ternary_Symbol, CDC}
- S = {IDS}, which is the start symbol of the grammar
- P: a set of rewrite rules

The following is the set of rewriting rules P:
- IDS::=<Binary_Symbol><IDS1><IDS1>|<Ternary_Symbol>
                <IDS1><IDS1><IDS1>
- <IDS1> ::= <IDS> | <CDC>
- <CDC>::= coded_ideograph | coded_radical | coded_component
- <Binary_Symbol> ::=⿰|⿱ | ⿴ | ⿵ | ⿶ | ⿷ | ⿸ | ⿹ | ⿺
- <Ternary_Symbol> ::= ⿲ | ⿳

Note that even though the IDCs are terminal symbols, they are not part of the Character Description Components.

M-set (main set): M-set is the set of characters that have been reviewed and accepted by IRG member bodies without pending questions in the current working set.

New Source: Any CJK source that is newly submitted by IRG member bodies which is not yet accepted by ISO/IEC 10646, thus is not present in Section 23 of ISO/IEC 10646. Member bodies may first submit their new source to the IRG for acceptance. Once accepted, the characters in that source can be accepted by the IRG for consideration for inclusion in future extensions.   The IRG will also submit the source to WG2 for approval and inclusion in Section 23 of ISO/IEC 10646.

Source: A reputable published document such as a dictionary, a standardization document, or a well published and widely read or referenced book which the IRG would consider as authoritative such that the characters in this source are considered reliable and stable for consideration of inclusion. A set of ISO/IEC 10646 accepted sources is listed in Section 23 of the ISO/IEC 10646 document.

Working set: A working set is the set of characters accepted by the IRG as a collection to work on for extension to ISO/IEC 10646. Characters accepted in a working set are subject to review by IRG member bodies for inclusion in a particular extension.

Deleted: ,

Deleted: 5