**ISO/IEC JTC 1/SC 2/WG 2**
**PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS**
**FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646**[1]
**Please fill all the sections A, B and C below.**
**Please read Principles and Procedures Document (P & P) from** http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and details before filling this form.**
**Please ensure you are using the latest Form from** http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html**.**
**See also** http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

### A. Administrative

1. **Title:** *Proposal to encode CJK Unified Ideographs Extension C1 in SIP*
2. Requester's name: *ISO/IEC JTC1/SC2/WG2/ IRG (Ideographic Rapporteur Group)*
3. Requester type (Member body/Liaison/Individual contribution): *Sub WG*
4. Submission date: *2006-08-27*
5. Requester's reference (if applicable):
6. Choose one of the following:
    This is a complete proposal: *Yes*
    (or) More information will be provided later:

### B. Technical – General

1. Choose one of the following:
    a. This proposal is for a new script (set of characters): *No*
       Proposed name of script:
    b. The proposal is for addition of character(s) to an existing block: *No*
       Name of the existing block:
2. Number of characters in proposal: *4,219*
3. Proposed category (select one from below - see section 2.2 of P&P document):
    A-Contemporary *X*   B.1-Specialized (small collection)   B.2-Specialized (large collection)
    C-Major extinct   D-Attested extinct   E-Minor extinct
    F-Archaic Hieroglyphic or Ideographic   G-Obscure or questionable usage symbols
4. Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document): *1*
    Is a rationale provided for the choice? *No*
       If Yes, reference:
5. Is a repertoire including character names provided? *No*
    a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?
    b. Are the character shapes attached in a legible form suitable for review? *Yes*
6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? *ISO/IEC JTC1/SC2/WG2/ IRG (Ideographic Rapporteur Group)*
    If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: *Mr. Chen Zhuang, Project Editor, CHENZH@cesi.ac.cn*
7. References:
    a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *Yes*
    b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? *Yes*
8. Special encoding issues:
    Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? *No*

9. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information.  See the Unicode standard at http://www.unicode.org for such information on other scripts.  Also see http://www.unicode.org/Public/UNIDATA/UCD.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

---

[1] Form number: N3002-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10)

**C. Technical - Justification**

1. Has this proposal for addition of character(s) been submitted before?  *No*
   If YES explain
2. Has contact been made to members of the user community (for example: National Body,
   user groups of the script or characters, other experts, etc.)?  *Yes*
       If YES, with whom?  *IRG member bodies*
       If YES, available relevant documents:
3. Information on the user community for the proposed characters (for example:
   size, demographics, information technology use, or publishing use) is included?  *Yes*
   Reference:  *Attachment*
4. The context of use for the proposed characters (type of use; common or rare)  *Common and rare*

   Reference:  *Attachment*
5. Are the proposed characters in current use by the user community?  *Yes*
       If YES, where?  Reference:  *Attachment*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely
   in the BMP?  *No*
           If YES, is a rationale provided?
           If YES, reference:
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?  *Yes*
8. Can any of the proposed characters be considered a presentation form of an existing
   character or character sequence?  *No*
           If YES, is a rationale for its inclusion provided?
           If YES, reference:
9. Can any of the proposed characters be encoded using a composed character sequence of either
   existing characters or other proposed characters?  *No*
           If YES, is a rationale for its inclusion provided?
           If YES, reference:
10. Can any of the proposed character(s) be considered to be similar (in appearance or function)
    to an existing character?  *No*
           If YES, is a rationale for its inclusion provided?
           If YES, reference:
11. Does the proposal include use of combining characters and/or use of composite sequences?  *No*
    If YES, is a rationale for such use provided?
           If YES, reference:
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
           If YES, reference:
12. Does the proposal contain characters with any special properties such as
    control function or similar semantics?  *No*
           If YES, describe in detail (include attachment if necessary)


13. Does the proposal contain any Ideographic compatibility character(s)?  *No*
    If YES, is the equivalent corresponding unified ideographic character(s) identified?
           If YES, reference:

Further Information

HISTORY:

The CJK Unified Ideographs Extension C1 project (CJK_C1) formally started in ISO/IEC JTC1/SC2/WG2/IRG in 2002 (IRG#19, Macao).  The initially proposed characters are over 20,000, coming from China, Japan, R. O. Korea, Vietnam, USA/Unicode, Hong Kong SAR, Macao SAR and TCA (Taipei Computer Association). In order to ensure quality and speed up the project, IRG members agreed to reduce the size of CJK_C1 to about 5,000 characters (IRG#25, Berkeley, 2005-12). The current version of CJK_C1 contains 4,219 characters; all of them had been reviewed by IRG members for more than 15 rounds.

STATISTICS:

There are 4,219 characters proposed to be encoded (See **Attachment A** for the multi-column listing).
● China: 1,127
● Japan: 369
● R. O. Korea: 405
● USA/Unicode: 77
● Macao SAR: 16
● TCA: 1,812
● Vietnam: 785
Note: some characters are contributed by 2 or more IRG members and are unified according to Ideograph Unification Rules. (See Annex S of ISO/IEC 10646)

The characters are selected from or used for:
● Dictionaries
● Sutra
● Other ancient or modern books
● Names of people
● Names of animals or plants
● Names of places
● Others

References:

For detailed information on the evidences of these characters provided by IRG members, you can access to the following documents under IRG Meeting 26(Hue, Vietnam) through the website: http://www.cse.cuhk.edu.hk/~irg.
● IRGN1225, Japan
● IRGN1227, China
● IRGN1228, Masao SAR
● IRGN1231, Vietnam
● IRGN1232, TCA
● IRGN1234, R. O. Korea
An attached excel file (**Attachment B**) gives more detailed source information of every character.