Title: Proposal to Add a Set of Compatibility Ideographs for Government Use
Date: 2008-10-10
Source: Japan

Japanese government has been working for years on comprehensive reformation of public administration, utilizing information technologies.   It is recognized that one of the barriers is so-called *gaiji* issue.

In some areas in government operations, there are some needs on handling huge amount of rare ideographs, which are primarily parts of proper nouns such as human names or place names.   Those ideographs not encoded in the coded character set standard are called *gaiji*.   Many *gaiji* are more or less variants of other ideographs.

Recently, a government funded group was formed and worked on the *gaiji* issue. Japan national body has a channel with the group.   The group researched three major administration systems and is compiling *gaiji* collected from those systems. The systems the group got input from are: *jumin kihon daicho* (住民基本台帳) network, *koseki* (戸籍) system, and new *toki* (登記) system.   They are all some sort of low-mandated government-operated registries.   Based on their preliminary reports, Japan national body submitted several ideographs to IRG, and we will some more in a future.   However, there are another set of *gaiji*, that is, a set of ideographs that are glyph variants of already encoded unified ideographs, although the government's operating rules mandate distinction.

Japan national body hereby proposes encoding of those ideographs as compatibility ideographs in ISO/IEC 10646.

List of attachments:
N3530-submission_form.doc   Summary form of this proposa, based on P&P.
n3530-mapping.txt   Mapping table between the proposed compatibility
                            ideographs and existing unified ideographs.
n3530-table_p01-20.pdf   List of proposed characters, part 1 of 4.
n3530-table_p21-40.pdf   List of proposed characters, part 2 of 4.
n3530-table_p41-60.pdf   List of proposed characters, part 3 of 4.
n3530-table_p61-80.pdf   List of proposed characters, part 4 of 4.

<table>
<tr><td colspan="2">

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*.**

</td></tr>
</table>

A. Administrative

1. **Title:**	Proposal to Add a Set of Compatibility Ideographs for Government Use
2. Requester's name:	*Japan National Body*
3. Requester type (Member body/Liaison/Individual contribution):	*Member body*
4. Submission date:	*October 10, 2008*
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	
(or) More information will be provided later:	*yes*

B. Technical – General

1. Choose one of the following:	
a. This proposal is for a new script (set of characters):	
Proposed name of script:	
b. The proposal is for addition of character(s) to an existing block:	*yes*
Name of the existing block:	*CJK Compatibility Ideographs*
2. Number of characters in proposal:	*2,621*
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   G-Obscure or questionable usage symbols

Ideographic

4. Is a repertoire including character names provided?	*yes*
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	*yes*
b. Are the character shapes attached in a legible form suitable for review?	*yes*

5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Japan

If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools

used:

6. 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?

no

7. 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)?

yes

CJK Compatibility Characters should be better handled by existing implementation on all aspects above as long as normalization process is not involved.

8. 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.

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? — *National Body*

        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?

    Reference: — *Online administration services which involve registration/reference of people's names and place names.*

4. The context of use for the proposed characters (type of use; common or rare) — *common*

    Reference:

5. Are the proposed characters in current use by the user community? — *yes*

    If YES, where? Reference:

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? — *yes*

        If YES, is a rationale for its inclusion provided? — *yes*

        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? — *yes*

        If YES, is a rationale for its inclusion provided? — *yes*

        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)?   *yes*

   If YES, is the equivalent corresponding unified ideographic character(s) identified?   *yes*

     If YES, reference: *Will be attached (n3530-mapping.txt).*

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 Add a Set of Compatibility Ideographs for Government Use
2. Requester's name: *Japan National Body*
3. Requester type (Member body/Liaison/Individual contribution): *Member body*
4. Submission date: *October 10, 2008*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal:
   (or) More information will be provided later: *yes*

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters):
      Proposed name of script:
   b. The proposal is for addition of character(s) to an existing block: *yes*
      Name of the existing block: *CJK Compatibility Ideographs*
2. Number of characters in proposal: *2,621*
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. Is a repertoire including character names provided? *yes*
   a. If YES, are the names in accordance with the "character naming guidelines"
      in Annex L of P&P document? *yes*
   b. Are the character shapes attached in a legible form suitable for review? *yes*
5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? *Japan*
   If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:
6. 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? *no*
7. 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)? *yes*
   CJK Compatibility Characters should be better handled by existing implementation on all aspects above as long as normalization process is not involved.

8. 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: N3152-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, 2007-03, 2008-05)

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?   *National Body*
            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?
      Reference: *Online administration services which involve registration/reference of people's names and place names.*
4. The context of use for the proposed characters (type of use; common or rare)   *common*
      Reference:
5. Are the proposed characters in current use by the user community?   *yes*
      If YES, where? Reference:
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?   *yes*
            If YES, is a rationale for its inclusion provided?   *yes*
            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?   *yes*
            If YES, is a rationale for its inclusion provided?   *yes*
            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)?   *yes*
      If YES, is the equivalent corresponding unified ideographic character(s) identified?   *yes*
            If YES, reference: *Will be attached (n3530-mapping.txt).*