

ISO/IEC JTC 1/SC 2/WG 2 N \_\_\_\_\_

ISO/IEC JTC 1/SC 2/WG 2/IRG N 1053

**ISO/IEC JTC1/SC2/WG2/IRG  
Ideographic Rapporteur Group  
(IRG)**

<b>TITLE:</b>	Request for Collection Identifier for International Ideograph Core Subset in ISO/IEC 10646
<b>SOURCE:</b>	IRG Rapporteur
<b>STATUS:</b>	Submission from the IRG to WG2
<b>DISTRIBUTION:</b>	Members of ISO/IEC JTC1/SC2/WG2
<b>DATE:</b>	2004-5-27
<b>REFERENCE:</b>	WG2 N2539 (IRG N947), IRG N1018

**Request For Collection Identifier**

**For a Sub-Repertoire Of ISO/IEC 10646**

<b>Date:</b>	2004.05.27
<b>SOURCE:</b>	<b>Ideographic Rapporteur Group</b>
<b>Email address of source:</b>	<a href="mailto:joe.zhang@unihan.com.cn">joe.zhang@unihan.com.cn</a>
<b>Phone number of source:</b>	_(86) 10 6484 9484Ext. 208_____
<b>Fax number of source:</b>	_(86) 10 6484 9487
<b>Address of source:</b>	<b>F06-29C, Forum, Datun, Chao Yang District, Beijing, China, zip 100101</b>
<b>WG 2 SPONSOR</b>	<b>Ideographic Rapporteur Group</b>
<b>SUBMITTER'S REFERENCE:</b>	<b>WG2 N2539 (IRG N947), IICore2.1.txt(IRG N1052)</b>

The IRG had the 22<sup>nd</sup> meeting in Chengdu, Sichuan, China, and succeeded in defining a stable International CJK basic subset in ISO/IEC 10646. WG2 N2539 (IRG N947) should address background information in detail on why the IRG has been engaging in the development of the sub-repertoire from CJK Unified Ideographs and its extensions in ISO/IEC 10646 to be identified

according to Annex A. The list of code positions for this collection is attached in the file IICORE2.1.txt(IRG N1052).

### **Specification of the proposed collection**

**Proposed Collection Name:** IICORE

**Number of characters:** 9,811

**Collection to be marked as Fixed (Yes / No):** YES

**Code positions in the Collection:** (See IICore2.1.txt)

### **Collections containing the proposed sub-repertoire:**

<b><i>ID</i></b>	<b><i>UCS-Collection Name / Code Positions</i></b>	<b><i>Positions to be included or excluded</i></b>
<u>60</u>	<u>CJK UNIFIED IDEOGRAPHHS 4E00-9FFF</u>	<u>Only 9,707 code positions specified in the file IICore2.1.txt are included.</u>
<u>81</u>	<u>CJK UNIFIED IDEOGRAPHHS EXTENSION A 3400-4DBF, FA1F, FA23</u>	<u>Only 42 code positions specified in the file IICore2.1.txt are included.</u>
<u>2001</u>	<u>CJK UNIFIED IDEOGRAPHHS EXTENSION B 20000-2A6DF</u>	<u>Only 62 code positions specified in the file IICore2.1.txt are included.</u>

### **Justification (In response to Annex E of WG2 N2652R):**

- A. IICORE collection is a large, sparse subrepartoire over three blocks. Without this collection, users need to enumerate all code positions of IICore. That is inconvenient and impractical.
- B. All characters in IICore 2.1 are defined in ISO/IEC 10646:2003.
- C. No existing collections are fully included in IICORE.
- D. Since no collection is fully contained in IICore, all code positions in IICore are listed in IICore2.1.txt.
- E. N/A.
- F. All code positions in IICORE are already defined.
- G. IICORE collection should be marked as FIXED, since all code positions are already defined.
- H. See WG2 N2539 (IRG N947) and IRG N1018

### **Organization of IICORE2.1.txt:**

1<sup>st</sup> column: the ISO/IEC 10646 code point.(0xxxx for BMP; 2xxxx for Plane 2)

2<sup>nd</sup> column: G source identifier

3<sup>rd</sup> column: T source identifier  
4<sup>th</sup> column: J source identifier  
5<sup>th</sup> column: H source identifier  
6<sup>th</sup> column: K source identifier  
7<sup>th</sup> column: M source identifier  
8<sup>th</sup> column: KP source identifier  
9<sup>th</sup> column: Category for IRG submission reference

**Note: the UCS code position on the first column is definitive, and other columns are informative for the purpose of the definition of IICORE collection and are explained below.**

#### **Source identifier description in 2<sup>nd</sup>-8<sup>th</sup> column**

Source identifier consists from 3 letters

The 1<sup>st</sup> letter G, T, J, H, K, M, P are member body identifiers which represent China, TCA, Japan, HKSAR, ROK, MacaoSAR, DPRK, respectively.

The 2<sup>nd</sup> letter followed by the member body identifier indicates the source subID, such as “0” for G-0, “7” for T-7, and so on. Sources which are not subdivided uses “1”.

The last letter in source identifier indicates one of the following, depending on where the character comes from:

- A: for level 1 of the source encoded character set
- B: for education
- C: for level 2 of the source encoded character set
- D: for personal names
- E: for place names
- F: for colloquial characters
- G: for anything else

#### **Category description in the 9<sup>th</sup> column**

**Category A characters:** Characters in the respective primary sets, such as the level one characters from GB2312 and JIS X 0208, of the source standards are referred to as Category A characters.

**Category B characters:** Other characters not in Category A with multiple sources are referred to as Category B characters.

**Category C characters:** Characters not in Category A with only a single source are referred to as Category C characters.

--- end of document ---