

ISO/IEC JTC1/SC2/WG2/IRG

Ideographic Rapporteur Group

(IRG)

Source/Contribution Identifier : IRG Rapporteur
Meeting : For WG 2 Meeting #45, Xia Men
Title : IICore Source use information
Status :

1. An extract of the IICore source identifier¹ data format *from the document IRGN1053* is recapped below in *italic* for easy reference.

Source identifier description in 2nd-8th column

Source identifier consists from 3 letters

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

The 2nd 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 9th 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.*

¹ as source use identifier according to WG2 meeting 45 terminology

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.

2. Members submitted the following elaboration to their respective Source use ID:

Submitter	Source ID and subID	Description
China	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
	G8	GB8565-88
	GE	GB16500-95
	G9	GB18030-2000
DPRK	P0	KPS 9566-97
Japan	J1A	IPSJ-TS 0007:2004 (Basic Subset of Coded Character Sets - Japanese Core Ideographs)
HKSAR	H1	The characters are from a) a publication widely used in primary education (<<常用字字形表>> in Chinese name only), b) Hong Kong Supplementary Character Set; and c) characters popularly used in names of people, places and companies as well as used in the local Cantonese dialect.
Korea	K0	KS0 KS X 1001
	K1	KS1 KS X 1002
	K2	KS2 KS X 1005-1
	K3	KS3 KS X 1005-2

Submitter	Source ID and subID	Description
Macau SAR	M1	<p>The main sources are:</p> <ul style="list-style-type: none"> • Newspaper: Macao Daily News and Va Kio Daily News, mainly from the year of 2000 to 2003. • Government departments: Identification Department for names of persons, Civic and Municipal Affairs Bureau for names of places. • Journals: Government magazines and journals. <p>Based on the information above, we have conducted a character frequency analysis. The range included is up to 99.94% accumulated frequency, with some adjustments of commonly used characters from names of persons and places in Macao.</p>
TCA	T1 T2 T3 T4 T5 T6 T7 TC TD TE TF	CNS 11643 plane 1 CNS 11643 plane 2 CNS 11643 plane 3 CNS 11643 plane 4 CNS 11643 plane 5 CNS 11643 plane 6 CNS 11643 plane 7 CNS 11643 plane 12 CNS 11643 plane 13 CNS 11643 plane 14 CNS 11643 plane 15

End of document

ISO/IEC JTC1/SC2/WG2/IRG

Ideographic Rapporteur Group

(IRG)

Source/Contribution Identifier : IICore Ad Hoc Group

Meeting : 23rd IRG Meeting in Cheju,
Republic of Korea

Title : IICore Ad Hoc Group Report

Status :

Notes of discussion:

1. The Group recapped that the IICore version 2.2 was accepted by the WG2 #45 meeting, and would be included in the ISO 10646:2003 Amendment 1.
2. The Group discussed the issue of source identifier data of the IICore. The need of maintaining source identifier data of ideographs in the IICore was reviewed. The Group agreed that the source identifier data is necessary, though they may not be required by the WG2.
3. It was agreed to keep the source identifier data by the IRG for future reference, whereas list of code position only would be maintained in the IICore under the ISO 10646:2003 Amendment 1.
4. The document management issue was discussed. (Ref. IRGN1089).
5. The source identifier data structure used by the China triggering proposed changes to 100 IICore characters (in IICore version 2.2) was reviewed.
6. China advised that the proposed changes to the source identifier data structure were for tracking of data from different standards in the China. China advised the Group that the proposed changes to 116 IICore characters (including 100 characters mentioned above in paragraph 5 and 16 new changes) would be stable

and no other changes in foreseeable future. The Group accepted China's clarification and agreed to accept the proposed changes.

7. The Group agreed that the IRGN1067 would have to be amended as a consequence of the acceptance to the above changes in paragraph 6 above. A revised version (IRGN1067R with source identifier data) will be prepared. The IRG Rapporteur will submit the IRGN1095 (with full list of IICore characters excluding source identifier data) to the ISO 10646 Project Editor.
8. The Group further requested Members to submit elaboration of the proper interpretation of the source identifier data, in particular the subID. Since DPRK is absent from this meeting, the IRG Rapporteur will request DPRK to elaborate on their source reference data (P0) before the next IRG Meeting.
9. An extract of the IICore source identifier data format *from the document IRGN1053* is recapped below in *italic* for easy reference.

Source identifier description in 2nd-8th column

Source identifier consists from 3 letters

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

The 2nd 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 9th 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.*

10. Members submitted the following elaboration to their respective Source ID:

Submitter	Source ID and subID	Description
China	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
	G8	GB8565-88
	GE	GB16500-95
	GX	GB18030-2000
DPRK	P0	To be contacted.
Japan	J1A	IPSJ-TS 0007:2004 (Basic Subset of Coded Character Sets - Japanese Core Ideographs)
HKSAR	H1	The characters are from a) a publication widely used in primary education (<<常用字字形表>> in Chinese name only), b) Hong Kong Supplementary Character Set; and c) characters popularly used in names of people, places and companies as well as used in the local Cantonese dialect.
Korea	K0	KS0 KS X 1001
	K1	KS1 KS X 1002
	K2	KS2 KS X 1005-1
	K3	KS3 KS X 1005-2

Submitter	Source ID and subID	Description
Macau SAR	M1	<p>The main sources are:</p> <ul style="list-style-type: none"> • Newspaper: Macao Daily News and Va Kio Daily News, mainly from the year of 2000 to 2003. • Government departments: Identification Department for names of persons, Civic and Municipal Affairs Bureau for names of places. • Journals: Government magazines and journals. <p>Based on the information above, we have conducted a character frequency analysis. The range included is up to 99.94% accumulated frequency, with some adjustments of commonly used characters from names of persons and places in Macao.</p>
TCA	T1 T2 T3 T4 T5 T6 T7 TC TD TE TF	CNS 11643 plane 1 CNS 11643 plane 2 CNS 11643 plane 3 CNS 11643 plane 4 CNS 11643 plane 5 CNS 11643 plane 6 CNS 11643 plane 7 CNS 11643 plane 12 CNS 11643 plane 13 CNS 11643 plane 14 CNS 11643 plane 15

End of document