ISO/IEC JTC1/SC2/WG2 IRG N2228

Date: 2018-3-15

Source:	China
Meeting:	IRG#50, Beijing, China
Title:	China's Submission of CJK_H Ideographs
Status:	Member's submission
Actions required:	To be considered by IRG
Distribution:	IRG
Medium:	Electronic
Pages:	15
Appendixes:	8

China's submission for CJK Unified Ideographs Extension H / IRG Working Set 2017 includes 949 characters. The ideographs were collected from the following sources.

Part One: 559 Chinese characters

- GDM-xxxxx：Place-name characters. Source from Public Order Administration，The Ministry of Public Security of the People's Republic of China.

- GXM-xxxxx：Human-name characters. Source from Public Order Administration，The Ministry of Public Security of the People's Republic of China.

- GKJ-xxxxx：Science and technology characters. Source from China National Committee for Terms in Sciences and Technologies.

    All these for WS2017 are Chinese medicine characters，which were collected from the following books：

    ZhongYao Dacidian《中药大辞典》，2006年，上海科学技术出版社，ISBN 7-5323-8271-0

    ZhongYi Zidian《中医字典》，1988年，河南科学技术出版社，J4245.83

    ZhongYi Yiji Zidian《中医医籍字典》，1989年，江西科学技术出版社，ISBN 7-5390-0151-8/R.27

    ZhongYi Jingdian Zidian《中医经典字典》，1990年，重庆出版社，ISBN 7-5366-0058-5

- GHC-xxxxx：HanYu Dacidian 《汉语大词典》，1986年，上海辞书出版社，9187.11 / 1988-1993年，汉语大词典出版社（found in

1988），ISBN 7-5432-0001-5/H.2

Part Two: 390 Zhuang characters
- GZ-xxxxxxx ：Ancient Zhuang Character Dictionary 《古壮字字典》：1989年广西民族出版社，ISBN 7-5363-0614-8，source reference G_Zxxxxxxx where the number part consists of page (3 digits) column (1 digit) entry (1 digit) character (2 digits).
- GPGLG-1xxx：Zhuang Folk Song Culture Series - Pingguo County Liao Songs - Long Songs 《壮族民歌文化丛书•平果嘹歌•长歌集》2004年广西民族出版社，ISBN 7-5363-4820-7。
- GPGLG-2xxx：Zhuang Folk Song Culture Series - Pingguo County Liao Songs - Love Songs 《壮族民歌文化丛书•平果嘹歌•恋歌集》2005年广西民族出版社，ISBN 7-5363-5014-7。
- GPGLG-3xxx：Zhuang Folk Song Culture Series - Pingguo County Liao Songs - Prose Songs 《壮族民歌文化丛书•平果嘹歌•散歌集》2005年广西民族出版社，ISBN 7-5363-5013-9。
- G_PGLG4xxx：Zhuang Folk Song Culture Series - Pingguo County Liao Songs - New Songs 《壮族民歌文化丛书•平果嘹歌•新歌集》2006年广西民族出版社，ISBN 7-5363-5012-0。
- GPGLG-5xxx：Zhuang Folk Song Culture Series - Pingguo County Liao Songs - Borrowed Songs 《壮族民歌文化丛书•平果嘹歌•客歌集》2005年广西民族出版社，ISBN 7-5363-5015-5。
- GLGYJ-xxxx: Zhuang Liao Songs Research 《壮族嘹歌研究》2008年广西民族出版社，ISBN 978-7-5363-5069-4。
- GZA-1xxxxx: A Vibrant and Unbroken Transmission - Filial Piety and Zhuang Funeral Songs 《生生不息的传承•孝与壮族行孝歌之研究》2010年北京民族出版社，ISBN 978-7-1051-0648-6。
- GZA-2xxxxx:《壮族伦理道德长诗传扬歌译注》2005年广西民族出版社，ISBN 7-5363-4922-X。
- GZA-3xxxxx: Compendium of Old Zhuang Folksong Texts - Wooing Songs vol. 1 – Liao Songs 《壮族民歌古籍集成•情歌（一）嘹歌》1993广西民族出版社，ISBN 7-5363-2714-5。
- GZA-4xxxxx: Compendium of Old Zhuang Folksong Texts - Wooing Songs vol. 2 – Fwen Nganx《壮族民歌古籍集成•情歌（二）欢榾》 1997年南宁广西民族出版社，ISBN 7-5363-3298-X。
- GZA-5xxxxx: Zhuang Character Database – Internet Collection《方块壮字资料库•网络集》http://yyysk.gxu.edu.cn/dspp/wangluoji.pdf 2011- 年广西大学 ，桂ICP备05000940号。
- GZA-6xxxxx：《中国壮族谚语》2015广州世界图书出版广东有限公司，ISBN 978-7-192-0492-1。
- GZA-7xxxxx: Ancient Remembrance - Zhuang Creation Myth Songs 《远古的追忆•壮族创世神话古歌研究》2012年北京民族出版社，ISBN 978-7-1051-2242-4。

Appendix 1: 559_HanZi_Glyphs
Appendix 2: 559_HanZi_Attributes
Appendix 3: 559_HanZi_Evidences
Appendix 4: 559_HanZi_Bitmaps
Appendix 5: 390_Zhuang_Glyphs
Appendix 6: 390_Zhuang_Attributes
Appendix 7: 390_Zhuang_Evidences
Appendix 8: 390_Zhuang_Bitmaps

Instruction of GDM & GXM

1. Source

The GDM & GXM for WS 2017 are all from Special character library for census information of public security, which is affiliated to The Ministry of Public Security of the People's Republic of China.

In the past 65 years, the Police have collected numerous uncommon Human-Name and Place-Name characters. Some local departments have already developed font library, while most of them still do the registration work by hand-writing.

In 2015, The Ministry of Public Security of the People's Republic of China entrusted Beijing Beida Fangzheng Electronic Co., Ltd. to upgrade the character library, with 1,639 characters delivered, to solve the problem that uncommon Human-Name and Place-Name can't be input into Ministerial population information system. In 2016, the actual delivery quantity is 3,014 of which 454 are Place-Name and 2,559 are Human-Name.

For the Confidentiality agreement signed between The Ministry of Public Security of the People's Republic of China and Beijing Beida Fangzheng Electronic Co., Ltd., some of the documentary evidences could not be published, only the following files are allowed to be used.

中华人民共和国公安部

公治办〔2015〕2089 号

关于提供人口信息冷僻字应用情况的函

北京北大方正电子有限公司：

　　为进一步升级完善公安人口信息专用字库，更好地服务广大人民群众，根据我局与你公司签署的《解决人口信息冷僻字问题合作协议》和 2010 年我局向你公司发出的委托函，我局对目前部级人口信息管理系统（二期）人口基本信息库中未包含在现行国家和国际标准中的冷僻字进行了统计整理，共计 1639 个，其中姓氏中使用冷僻字 476 个，名字中使用 1610 个，地名中使用 466 个，曾用名中使用 1396 个（详情见随附电子文档）。现提供你公司，请进行进一步整理、汇总和确认工作，同时，建议纳入你公司参与的《国家"十一五"时期文化发展纲要》重大建设项目——"中华字库"，并逐步提请申报国家标准。

<div align="right">

公安部治安管理局

2015 年 9 月 14 日

</div>

上海
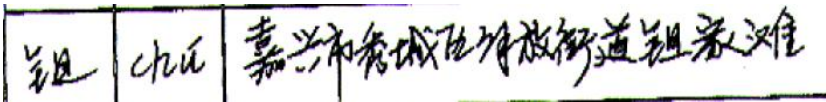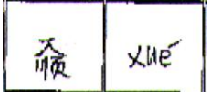
附件一：

地名冷僻字填报表　2个

部首	字形	拼音	所属市、县（区）、乡（镇）、村
王	璐	Lu	上海市崇明县港沿镇鲁东村
亻	俤	ge	民族

2. Selection principles
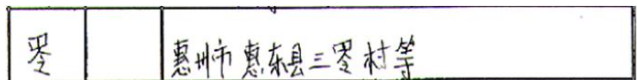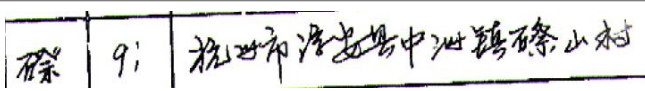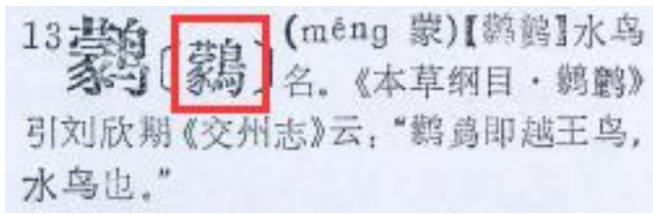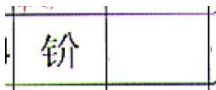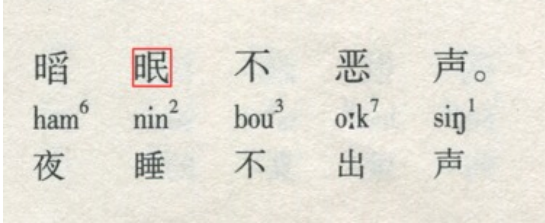
1) Clear evidence
The unclear images couldn't be selected:

| IDS	Evidence
⿰钅且	钮 | ⿰钅且 | 嘉兴市秀城区建设街道钮家滩
⿱大顺	顙 | xué
⿻门⿻关占	阘

2) No obvious error
The wrongly written characters and the ones with minor stroke differences shouldn't be selected:

IDS	Evidence	Correct glyph
⿻价令	罖 | 惠州市惠东县三罖村等	罖
⿰石④	磜 | 9ʲ | 杭州市淳安县中洲磜碥山村	磜
⿱蒙鳥	13 鸏 [鸏] (méng 蒙)【鹲鸏】水鸟 名。《本草纲目·鸏鹲》 引刘欣期《交州志》云，"鸏鸏即越王鸟， 水鸟也。"	鸏
⿱宀总	寁 | 汉族 | 福建省大田县 | 3	寁
⿰钅介	钐	钐
⿰氵乿	瀹 | 汉族 | 江苏省宜兴市 | 3	瀹

IDS	Evidence	Correct glyph
□日民	 暗 眠 不 恶 声。 ham⁶ nin² bou³ oːk⁷ siŋ¹ 夜 睡 不 出 声	眠

3) Can't be added to IVD

Target glyph	Evidence		
尭	尭	陕西省西安市临潼区	秦陵办王尭村庞沟组	1
叡	叡 rui		
桑	桑	汉族	福建省明溪县	3
敄	敄	汉族	江苏省江阴市	3

4) Respecting cultural inheritance

Hand-writing of grass-roots society is volatile, for convenience, or may be other reasons, caused the result that so many variant characters were created.

For the using of a variant glyph, which we can call Folk Characters, is often idiomatic, although most of the variants with minor stroke differences could be unified with the standardized form, some of the Place-Name characters which have been used for a long time should be maintained:

Source references	Bmp	IDS	analysis of character pattern	Code	Coded character

Source references	Bmp	IDS	analysis of character pattern	Code	Coded character
GDM-00064		□山领	Inexhaustive simplified of 嶺	E1DF	?
GDM-00078		□木🔢	Wrongly simplified of 樑	E1EA	?
GDM-00117		□石㘛	Ancient glyph of 砲	E200	?
GDM-00134		□纟厘	Wrongly simplified of 纏	E210	?
GDM-00152		□钅录	Wrongly simplified of 錄	E21E	?
GDM-00220		□纟宁	Wrongly simplified of 纻	E23D	?

The folk and uncommon characters was also constantly being used to name-child, then using simplified characters became a historical trend in the last 500 years, that's why we found so many half or wrongly simplified characters in the source from The Ministry of Public Security of the People's Republic of China. It's certainly getting their owners into trouble with the spread of digital and Internet technology, and it's believed that most of the owners were born before *The Law of Chinese Current Spoken and Written Language,* which provides the provisions of using characters for Human-name. It's nearly impossible for them to change their names into standardized form now, and as a specific phenomenon of a special period, some of this kind of characters can be recommended for Chinese character coding.

Source references	Bmp	IDS	analysis of character pattern	Code	Coded character
GXM-00014		□讠宁	Wrongly simplified of 訂	E24A	?
GXM-00071		□显页	Irregular simplified of 顕	E08C	?
GXM-00142		□舟风	Wrongly simplified of 艑	E0B4	?

Source references	Bmp	IDS	analysis of character pattern	Code	Coded character
GXM-00074		□钅监	Irregular simplified of 鑑	E08F	?
GXM-00035		□氵□ 门东	Wrongly simplified of 瀾	E06E	?
GXM-00165		□禾责	Irregular simplified of 積	E0BF	?
GXM-00189		□𧾷责	Irregular simplified of 蹟	E0CE	?
GXM-00208		□门豈	Wrongly simplified of 闓	E0D9	?
GXM-00372		□是刀	Possible variant of 剭	E162	?

Finally, we are very grateful to the young experts for their work and made a valuable contribution to the revision of the proposal.

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/irg31/IRGN1562.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/SubmissionForm.pdf **.**
See also http://appsrv.cse.cuhk.edu.hk/~irg/UCV.html **for latest** *Unifiable Calligraphic Variations*.

A. Administrative

1. **IRG Project Code:**	CJK Unified Ideographs Extension-H / IRG Working Set 2017
2. **Title:**	China's Submission of CJK_H Ideographs for 2017
3. Requester's region/country name:	China
4. Requester type (National Body/Individual contribution):	Member Body
5. Submission date:	2018-3-15
6. Requested Ideograph Type (Unified or Compatibility Ideograph	Unified Ideographs
If Compatibility, does requester have the intention to register them as IVS (See UT the IRG's approval? (Registration fee will not be charged if authorized by the IRG.)	
7. Request Type (Normal Request or Urgently Needed)	Normal Request
8. Choose one of the following:	
This is a complete proposal:	Yes
(or) More information will be provided later:	

B. Technical – General

1. Number of ideographs in the proposal:	*949*
2. Glyph format of the proposed ideographs: (128x128 "bmp" files or TrueType font file)	Bmp & TTF
If 'bmp' files, their file names are the same as their Source IDs?	*Yes*
If TrueType font, all proposed glyphs are put into BMP PUA area?	*Yes*
If TrueType font, data for Source IDs vs. character codes are provided?	*Yes*
3. Source IDs:	
Do all the proposed ideographs have a unique, proper Source ID (country/region o than 9 alphanumeric characters)?	Yes
4. Evidence:	
a. Do all the proposed ideographs have the separate evidence document which cor one scanned image of printed materials (preferably dictionaries)?	Yes
b. Do all the printed materials used for evidence provide enough information to tra third party (ISBN numbers, etc.)?	Yes
5. Attribute Data Format: (Excel file or CSV)	*Excel*

C. Technical - Checklist

Understandings of the Unification Checklist

1.	Has the requester read ISO/IEC 10646 Annex S and did the requester understand the uni policy?	*Yes*
2.	Has the requester read the "Unifiable Calligraphic Variations" (contact IRG technical ed the Rapporteurfor the latest one) and did the requester understand the unifiable variation	*Yes*
3.	Has the requester read this P&P document and did the requester understand the 5% rule?	*Yes*

Character-Glyph Duplication Checklist(http://www.itscj.ipsj.or.jp/sc2/open/pow.h **all the published ones and those under ballot)**

4.	Has the requester checked that any of the proposed ideographs is *not unifiable* with the compatibility ideographs of ISO/IEC 10646?	*Yes*
	If yes, which version of ISO/IEC 10646 did requester check? (e.g. 10646:2003)	10646: 2016 CD
5.	Has the requester checked that any of the proposed ideographs is *not unifiable* with the Amendments of current ISO/IEC 10646? (As of 2009, Amendment 1, 4, 5, 6and 8 have ideographs.)	*Yes*
	If yes, which amendments did requester check?	
6.	Has the requester checked that any of the proposed ideographs is *not unifiable* with the the current IRG working sets or proposed amendments of ISO/IEC 10646? (As of 2009, PDAM 8 have CJK ideographs.)	*Yes*
	If yes, which draft amendments did requester check?	
7.	Has the requester checked that any of the proposed ideographs is *not unifiable* with the the current working M-set and D-set of the IRG? (Contact IRG chief editor and technica through the IRG Rapporteur for the newest list)	*Yes*
	If yes, which document did requester check?	Extension G
8.	Has the requester checked that any of the proposed ideographs is *not unifiable* with the or mis-unified ideographs in ISO/IEC 10646? (Check Annex E of this document).	*Yes*
9.	Has the requester checked that any of the proposed ideographs *has similar ideograph(s)* ideographs in the current standardized or working set mentioned above?	*Yes*
10.	Has the requester checked that any of the proposed ideographs *has variant ideograph(s)* ideographs in the current standardized or working set mentioned above?	*Yes*

Attribute Data Checklist

11.	Do all the proposed ideographs have attribute data such as the KangXi radical code, stro first stroke?	*Yes*
12.	Are there any simplified ideographs (ideographs that are based on the policy described i in the proposed ideographs?	*Yes*
	If YES, does your proposal include proper simplified/traditional indication flag for e ideograph in attribute data?	*Yes*
13.	Do all the proposed ideographs have the document page number of evidence documents data?	*Yes*
14.	Do all the proposed ideographs have the proper Ideographic Description Sequence (IDS data?	*Yes*
	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 variant ideographs for the proposed ideographs?	*No*