Title: Draft GB 18030-2022 Amendment Feedback & Recommendations
To: China Electronics Standardization Institute (CESI)
From: United States National Body (USNB)
Date: 2023-02-03

Experts from the USNB carefully reviewed the draft GB 18030-2022 amendment, and discussed its ramifications. This document provides detailed feedback, along with a number of recommendations.

For the purposes of this document, all references to the ISO/IEC 10646 standard equally apply to the GB/T 13000 standard and the Unicode Standard.

Additional Characters in CJK Unified Ideographs Blocks

The 26 additional characters that correspond to—and fill—the first two CJK Unified Ideographs blocks, specifically 16 characters in the CJK Unified Ideographs block (9FF0..9FFF) and 10 characters in the CJK Unified Ideographs Extension A block (4DB6..4DBF), which will be within the scope of GB 18030-2022 Implementation Level 1, are a welcome—and predictable—addition.

The 14 additional characters that correspond to the CJK Unified Ideographs Extension B block (9 characters: 2A6D7..2A6DF) and the CJK Unified Ideographs Extension C block (5 characters: 2B735..2B739) are also a welcome addition, and help to align GB 18030-2022 with the ISO/IEC 10646 standard. The same is true of the addition of the CJK Unified Ideographs Extension G and CJK Unified Ideographs Extension H blocks. Furthermore, the proposed new implementation deadline for new characters in this draft amendment, 2024-08-01, gives vendors that support GB 18030-2022 Implementation Level 3 sufficient time to design the glyphs for the 9,131 characters in these two new CJK Unified Ideographs blocks.

Additional Characters in Plane 10

The 897 characters—described as 公安人口信息专用字库补充汉字 in the draft amendment—that will be encoded at the beginning of Plane 10 (A0000..A0380) are problematic in that they destabilize the synchronization with the ISO/IEC 10646 standard, because Plane 10 is reserved and unallocated in that standard. As a result, *it will be impossible* for any products that conform to ISO/IEC 10646 to also conform to such an amendment of GB 18030-2022.

Furthermore, some of the characters that will be encoded in Plane 10 are either 1) already encoded in a CJK Unified Ideographs block; 2) unifiable with characters in an existing CJK Unified Ideographs block per IRG (*Ideographic Research Group*) unification rules; or 3) already in the pipeline for encoding in the CJK Unified Ideographs Extension I block (aka IRG Working Set 2021).

An example for the first issue is 0xBE389037/U+A000D 圮心, which is already encoded as U+2BB7B 愻 in the CJK Unified Ideographs Extension E block. See the excerpts below:
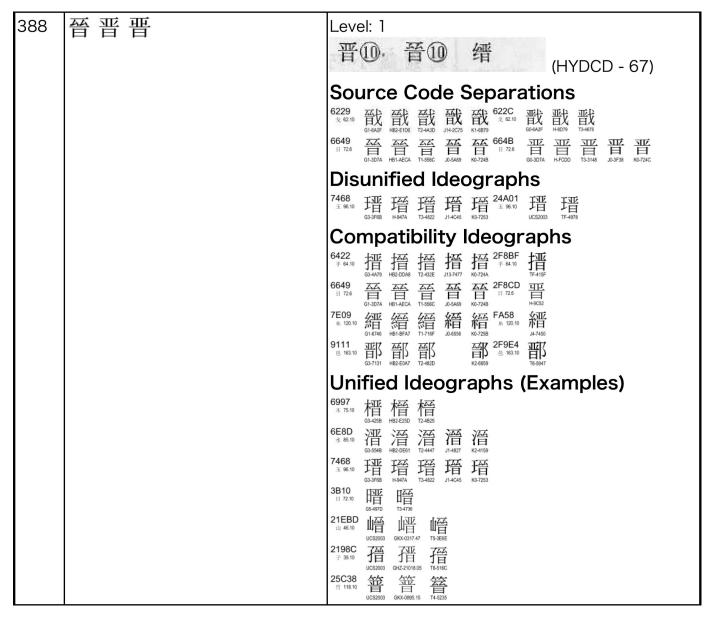
	30	31	32	33	34	35	36	37	38	39
8F					礋 A0000	奵 A0001	罙 A0002	圵 A0003	塄 A0004	塇 A0005
90	鏧 A0006	嘴 A0007	壌 A0008	卿 A0009	婷 A000A	崲 A000B	崔 A000C	恝 A000D	或 A000E	趴 A000F

2BB7B

土 32.7

恝

V4-445A

An example of the second issue is 0xBE38E839/U+A037F ☐玊 晉, which is unifiable with both U+7468 瑧/瑨 in the CJK Unified Ideographs block and U+24A01 瑨 in the CJK Unified Ideographs Extension B block per UCV (*Unifiable Component Variations*) #388 (see page 156 of IRG N2584). See the excerpts below:

	30	31	32	33	34	35	36	37	38	39
E3	筄 A0344	筃 A0345	筐 A0346	籔 A0347	羘 A0348	翹 A0349	緊 A034A	朢 A034B	魠 A034C	鏊 A034D
E4	餃 A034E	嗣 A034F	羢 A0350	岡 A0351	瓅 A0352	搖 A0353	菺 A0354	滢 A0355	梭 A0356	喞 A0357
E5	跌 A0358	蚚 A0359	鎑 A035A	锓 A035B	鑭 A035C	鎧 A035D	篓 A035E	儒 A035F	魅 A0360	魁 A0361
E6	顁 A0362	龟 A0363	拏 A0364	泉 A0365	悥 A0366	燦 A0367	瀚 A0368	彴 A0369	嫶 A036A	忍 A036B
E7	綢 A036C	鞻 A036D	勉 A036E	酏 A036F	缀 A0370	淩 A0371	顜 A0372	圳 A0373	驸 A0374	骒 A0375
E8	儊 A0376	瀞 A0377	瑩 A0378	� A0379	勇 A037A	烔 A037B	瑶 A037C	铮 A037D	谂 A037E	瑨 A037F
E9	燠 A0380									

7468

玉 96.10

瑨　瑨　瑨　瑨　瑨

G3-3F6B　H-947A　T3-4822　J1-4C45　K0-7253

24A01

玉 96.10

瑨

TF-4978

388	晉 晋 晋	Level: 1 (HYDCD - 67)

Source Code Separations

Disunified Ideographs

Compatibility Ideographs

Unified Ideographs (Examples)

An example of the third issue is 0xBE388F38/U+A0004 ▯土黽, which is IRG Working Set 2021 Serial Number #00833 that was submitted by China as GDM-00285. See the excerpts below:

BE38

8F	30	31	32	33	34	35	36	37	38	39
					㲝	歼	宷	𡈽	堩	㘃
					A0000	A0001	A0002	A0003	A0004	A0005

00833	土 32.8.2	堩
	▯土黽	GDM-00285
	| (2) 11	

In addition, the characters that will be included in Plane 10 are not ordered as a single repertoire according to their indexing radicals, but seemingly as several smaller repertoires with no distinct break or gap between them.

Recommendations

The USNB formally makes the following recommendations:

- Remove altogether from this draft amendment the repertoire of 897 characters in Plane 10 in order to preserve synchronization with the ISO/IEC 10646 standard. For implementations that have an immediate need to support this repertoire of 897 characters, they should be mapped to code points that correspond to one of the three PUA (*Private Use Area*) blocks. Furthermore, these PUA characters should not be documented in this amendment. After more than 17 years, GB 18030-2022 finally removed the PUA requirement from this standard, so establishing a new PUA requirement, particularly for a large number of characters, is fundamentally a step backwards.
- Remove from the repertoire of 897 characters any character that is already encoded.
- Remove from the repertoire of 897 characters any character that is unifiable with an existing character in a CJK Unified Ideographs block according to IRG rules. These characters can be represented in "plain text" as registered IVSes (*Ideographic Variation Sequences*) in a new IVD (*Ideographic Variation Database*) collection according to the procedures described in UTS #37, *Unicode Ideographic Variation Database*.
- Remove from the repertoire of 897 characters any character that is currently in the pipeline for encoding, meaning that the character is in IRG Working Set 2021.
- Submit to the IRG as a UNC (*Urgently Needed Character*) repertoire according to IRG Principles and Procedures (aka IRG N2515) all remaining characters in the repertoire of 897 characters.

That is all.