Title:	USNB Comments on Draft 2 of GB 18030-2022 Amendment 1 and recommendation for ISO/IEC 10646:2020 Amendment 2
Source:	US
Status:	National Body Contribution
Action:	For consideration by SC2 and China NB
Date:	2023-05-01

Background

China is preparing an amendment of their GB 18030-2022 national standard, "GB 18030—2022《信息技术 中文编码字符集》" ('GB 18030: 2022 Information Technology—Chinese Coded Character Set'). A major portion of that standard specifies CJK ideographs for which support is required. These characters are documented in terms of a legacy, multibyte encoding and also in terms of corresponding code points in ISO/IEC 10646 ("UCS"). Thus, the details of that specification are subject to the conformance requirements of the international standard and, hence, the development of GB 18030-2022 is of direct relevance to JTC 1/SC 2.

A first draft of the GB 18030-2022 amendment was made publicly available for comments late in 2022: "１号修改单 （征求意见稿）" ('Amendment 1 (Draft for solicitation of comments').[1] The US NB prepared comments for consideration by China, and these were also circulated to SC 2 because of potential implications for UCS.[2] In particular, the draft proposed to encode 897 CJK ideograph characters to code positions in plane 0A. (In a second draft, this number was reduced to 614. [3]) The US comments on the first draft stated,

> 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.
> …
> 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….[4]

---

[1] See L2/23-113.

[2] See L2/23-057. = SC 2/N4834.

[3] "１号修改单 （第二次征求意见稿）" ('Amendment 1 (Second Draft for Comments'). See L2/23-100.

[4] The repertoire of additional characters is described in the draft amendment as "公安人口信息专用字库补充汉字" ('Special Font Database for Public Security Information Supplemental Chinese Characters').

The US was not alone in commenting on this aspect of the draft.[5] However, in the published disposition of comments, this was rejected by China. Quoting from their disposition (with emphasis added in the English translation):[6]

> 不采纳。
> 目前 Unicode 标准与国际标准 ISO/IEC 10646 并未指定 0A 平面的用途。在此情况下，我国国家标准使用该空闲平面进行相关汉字编码并无不妥。

> Not adopted.
> The current Unicode standard and the international standard ISO/IEC 10646 do not specify the purpose of the 0A plane. In this case, *it is not inappropriate for China's national standard to use this idle plane* for relevant Chinese character encoding.

ISO/IEC 10646 conformance

It appears there may be some misunderstanding regarding the purpose of plane 0A and of the conformance requirements of the UCS:[7] Clause 6 clearly identifies plane 0A as reserved:

> **6 General structure of the UCS**
> …
> Planes from 04 to 0D are reserved for future standardization.

Then, later in that document, clause 7.3.8 clearly constrains any usage of reserved code points:

> **7.3.8 Reserved code points**
> Reserved code points are reserved for future standardization and shall not be used for any other purpose. Future editions of this document may allocate characters to some of these reserved code points.

These conformance requirements are reflected in China's national adoption of ISO/IEC 10646:2003, which is GB/T 13000:2010. The clause in ISO/IEC 10646:2003 that corresponds to clause 6 of the 6th edition is clause 9.1:

> **9.1 Planes accessible by UTF-16**
> …
> The Planes from 03 to 0D of Group 00 are reserved for future standardization.

The corresponding text in GB/T 13000:2010:

---

[5] Similar comments were also submitted by: "招商银行-纪熙东" ('Ji Xidong (China Merchants Bank)'); "个人专家-王谢杨" ('Wang Xieyang (Individual Expert)'). See footnote 6.

[6] "1 号修改单（第一次征求意见稿）意见汇总处理表" ('Amendment 1 (First Draft for Comments) Comments Summary Processing Form'). See L2/23-100.

[7] ISO/IEC 10646:2020.

9.1 UTF-16 可表示的平面

…

00 组的 03 到 0D 平面为未来的标准化保留。

The clause in ISO/IEC 10646:2003 that corresponds to clause 7.8.3 of the 6ᵗʰ edition is found in point b in clause 7:

7 General requirements for the UCS

…

b. Code positions to which a character is not allocated, except for the positions reserved for private use characters or for transformation formats, are reserved for future standardization and shall not be used for any other purpose.

The corresponding text in GB/T 13000:2010:

7 UCS 的总体要求

…

b. 除了那些为专用字符或传输格式分配的代码位置外，本标准中任何未安排字符的代码位置都应留待以后进一步标准化使用，而不得用于任何其他目的。

Plane 0A is not simply idle but is *reserved*, and usage for any purpose other than future standardization by JTC 1/SC 2 is not permitted. This prohibits use of reserved code points for any character data interchange. Standardization of characters to reserved code points requires the normal balloting processes of JTC 1/SC 2.

Recommendation

In consideration of this constraint in the UCS, yet also understanding China's urgent need for encoding of 600+ characters from their "special font database", experts in the Unicode Consortium have prepared a proposal for encoding of 603 CJK characters as *CJK Unified Ideographs Extension I* in available code positions within plane 02. (See L2/23-114 = SC2/N4841.) **The USNB endorses this proposal and *strongly recommends to SC2* that these characters be assigned and included in Amendment 2 of the 6ᵗʰ edition (ISO/IEC 10646:2020).** This will enable China to proceed in development of their national standard in conformance with the UCS.