

Title: Proposal to encode Small Seal Script in UCS

Source: TCA and China

Action: For consideration by JTC1/SC2/WG2

This document proposes to encode 11,108 Small Seal Script characters of Shuowen Jiezi (Tenghuaxie Version, 14 volumes in total) in UCS. Among those characters, there are 800 small Seal Script characters have been introduced on WG2#63 meeting. The appendix contains 14 parts, providing original glyphs, TTF glyphs, SW Radicals, Corresp Modern Chars. For WG2 discussion and feedback.

1 Introduction

The “Small Seal Script (小篆)” was resulted from the first standardization work for unifying Chinese characters of different ancient scripts by the government of Qin Dynasty (秦朝, 221 BC to 207 BC) of China. The Small Seal Script characters are known as the predecessors of modern Hanzi (漢字, i.e., Chinese characters). The first Chinese dictionary “Shuowen Jiezi (說文解字)”, edited by Xu Shen (許慎, 58 AD to 147 AD), collected most of the Small Seal Script characters from documents of Qin Dynasty and early Han Dynasty (漢朝, 202 BC to 220 AD). This dictionary included 9,353 canonical Small Seal Script characters and 1,163 variants. In book Shuowen Jiezi, those variants were called Chongwen (重文, i.e., duplicated characters), which included Zhouwen (籀文) and Guwen (古文). The Zhouwen was the Large Seal Script (大篆) used in Zhou Dynasty (周, 1,046 BC to 256 BC) and the Guwen were the scripts used in the other six nations excluding Qin in the Warring States period (戰國時期, 475 BC to 221 BC), which were incompatible with Small Seal characters.

There were two types of publicly accepted scripts prevailing in Qin Dynasty, canonical Small Seal Script and Clerical Script (隸書). The canonical Small Seal Script was used by government and for official documents, and the Clerical Script was used by the public. The Clerical Script was known as Qin Li (秦隸) in philology, and it was sometimes named Gu Li (古隸) in order to distinguish from subsequent Han Li (漢隸). Some people think Small Seal Script is a calligraphy typeface of modern Hanzi, which is a misunderstanding. In the book Shuowen Jiezi, the Small Seal Script is considered as a Chinese character set of a historical stage in the development of ancient Chinese characters, like Oracle Bone script (甲骨文), Bronze inscriptions (金文) and Warring States scripts (戰國文字), it does not belong to the same system as the modern Hanzi does. In addition, a large number of Small Seal Script characters cannot be mapped to modern standard Hanzi characters in form of shapes one by one; although some shapes can be mapped, the Small Seal Script characters have unique structures different to their counterparts in modern Hanzi. The Small Seal Script shall not be regarded as a typeface of modern Hanzi pertaining to standard Hanzi script, thus it should be encoded independently in UCS.

The Small Seal Script in the book Shuowen Jiezi plays an important role in the development of Hanzi, and gives a distinctive value to the study and collating of ancient characters:

(a) The Small Seal Script of Qin Dynasty is the final stage of the development of ancient

Hanzi characters, covering a complete structure model of pre-Qin ancient characters since Oracle bone inscriptions, and relates to the main components of the ancient Hanzi characters, and had become a required reference Chinese character set for doing textual research on and collating ancient Hanzi characters.

- (b) The Small Seal Script characters in the book *Shuowen Jiezi* are closely related to Zhou classic literatures of Zhou Dynasty, and the configuration of the characters can be explained by the meanings of the classic literatures; not only does it show the feature of the ideographic Chinese characters, but it also provides the method of composing modern Hanzi characters since the period of official scripts. Chinese characters are interpreted in accordance with the *Shuowen Jiezi* in current teaching.
- (c) Having strictly collated and removed unnecessary glyphs, the Small Seal Script in the book *Shuowen Jiezi* built a complete system of configuration. The characters are distributed in sequence according to their sub classes, under 540 main classes of radicals. Currently, ancient Hanzi characters are collated on the basis of the radical and character order of the book *Shuowen Jiezi*.
- (d) The 540 radicals summarized in the book *Shuowen Jiezi* cover all forms and meanings of Chinese characters, are called structural radicals with theoretical values. The 214 radicals of *Kanxi Zidian* (康熙字典, Kangxi Dictionary) of Qing Dynasty (清朝, 1644 to 1911) were created on the basis of simplifying and combining those 540 radicals. Such combination ignores the distinguishing feature of the united form and meaning of ideographic characters, they are merely suitable for a checking purpose, and are theoretically inadequate, the 540-radical is a more complete radical system.

2 Selection of Original Materials

The book *Shuowen Jiezi* was completed in Eastern Han Dynasty (東漢, 25 AD to 220 AD). After a long period of circulation, some of the original contents were changed. In Song Dynasty (宋朝, 960 AD to 1279 AD), Xu Xuan (徐鉉) was ordered by the Emperor to make a complete collation of the publication. He restored the contents and got its original glyphs. This book is known as “Daxu Ben (大徐本, the version of Big Xu)” while the book collated by his younger brother Xu Kai (徐鍇) is known as “Xiaoxu Ben (小徐本, the version of Little Xu)”. The “Daxu Ben” *Shuowen Jiezi* is recognized as a specification version. It was followed as major and principal reference when character glyphs were selected and ordered for this proposal.

There are 3 most popular versions of Daxu Ben *Shuowen Jiezi*: Tenghuaxie version (額勒布藤花榭本), Pingjinguang version (孫星衍平津館本) and Chenchangzhi version (陳昌治單行本, revised in accordance with Pingjinguang version). Both Tenghuaxie version and Pingjinguang version were made by following original Song Dynasty printed books, while Pingjinguang version was revised by the Chenchangzhi version. To retain the original contents as much as possible, Tenghuaxie version was selected first when the proposal was drafted.

Duan Yucai (段玉裁), a leading expert for the *Shuowen Jiezi* research in Qing Dynasty, collated Daxu Ben and Xiaoxu Ben and then corrected some glyphs according to the examples of the *Shuowen Jiezi* itself and adjusted partial ordering of characters. His work is reflected in the book *Shuowen Jiezi Zhu* (說文解字注, annotations for the book *Shuowen Jiezi*, abbreviated as Duan Zhu (段注)). The Duan Zhu is an important basis for people now to learn the *Shuowen Jiezi*. The different information between Daxu Ben and Duan Zhu was preserved based on Daxu Ben when the proposal was drafted.

3 Rules and Principles

3.1 Ordering

- (a) Characters are ordered based on the Shuowen Jiezi radical sequence and the arrangement order of the characters within the same radical partition of the Tenghuaxie version.
- (b) Characters from Duan Zhu or Chenchangzhi version are ordered as follows:
 - (1) Additional variant characters of each version are placed in order, which behind the corresponding Zhengzhuan (正篆) and Variants (重文).
 - (2) Newly added characters(新附)are placed at the end of other characters with the same radicals.

3.2 Selection of character

- (a) All Zhengzhuan (正篆) characters, Variants (重文) and newly added characters of Tenghuaxie version are to be included.
- (b) Characters that have different structures in Duan Zhu and Chenchangzhi version are to be supplemented. (Such as “蓋”)

3.3 Shapes of characters

Characters from Tenghuaxie version shall have shapes consistent with their original forms in the Tenghuaxie version. Added characters and component variants shall have shapes consistent with their original forms. No purposed adjustment or re-construction.

3.4 Principles for transferring Old Hanzi to modern Hanzi

- (a) Use descriptive glyph as the first corresponding glyph.
 - (1) Follow the original component position, e.g. 焮—焮；
 - (2) Follow the original conversion of line to stroke, e. g. 𠃉—𠃉
 - (3) Follow the original form of Li glyph, e. g. 𠃉—𠃉
- (b) If there is a corresponding natural script-evolution standard script, it should be used as a secondary corresponding glyph ° For example,

焮—焮—秋； 𠃉—𠃉—曲 Guwen； 𠃉—𠃉—中 Zhouwen

4 Format of the attached excel file

The attached excel file contains totally 11,116 entries, each for one Small Seal Script character, and ordered by the rules mentioned in clause 3.1. In the excel file, each entry consists 7 columns described as follows:

- (1) Id code: a unique number used to identify all corresponding image and/or description in other files,
- (2) THX's sequence no.: the sequence number in the Tenghuaxie version for the character,

- (3) THX's image: the image of the character which obtained from the Tenghuaxie version,
- (4) dot matrix: the reproduced dot matrix image of the character, according to the column “THX's image”,
- (5) corresponding Hanzi: the corresponding modern Chinese character of the character,
- (6) radical: the Shuowen Jiezi radical of the character, according to the book Tenghuaxie version, and
- (7) radical code: the sequence number of the Shuowen Jiezi radical.

5 Experts of the working group

TCA: 李鑾(Xian Li), 蔡信發(Xin-fa Tsai), 季旭昇(Hsiu-sheng Chi), 許學仁(Xue-jen Hsu), 宋建華(Chien-hua), 袁國華(Kwok-wa Yuen), 魏林梅(Lin-mei Wei), 莊素琴(Suh-chyin Chuang)

China: 王寧(Wang Ning), 王立軍(Wang Liun), 王曉明(Wang Xiaoming), 胡佳佳(Hu Jiajia), 師霞(Bu Shixia), 凌麗君(Ling Lijun), 孟琢(Meng Zhuo), 董婧辰(Dong Jingchen)

Organization: Chinese Foundation for Digitization Technology, the Project of Chinese Character Repertoire, China Electronics Standardization Institute

ISO/IEC JTC 1/SC 2/WG 2

PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest *Roadmaps*.

A. Administrative

1. Title:	Proposal to encode Small Seal Script in UCS
2. Requester's name:	TCA and China
3. Requester type (Member body/Liaison/Individual contribution):	Liaison and Member body
4. Submission date:	2015-10-20
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal: _____	Yes
(or) More information will be provided later: _____	

B. Technical – General

1. Choose one of the following:					
a. This proposal is for a new script (set of characters):	Yes				
Proposed name of script:	Small Seal Script				
b. The proposal is for addition of character(s) to an existing block:	No				
Name of the existing block:					
2. Number of characters in proposal:	11,108				
3. Proposed category (select one from below - see section 2.2 of P&P document):	F				
A-Contemporary	<input type="checkbox"/>	B.1-Specialized (small collection)	<input type="checkbox"/>	B.2-Specialized (large collection)	<input type="checkbox"/>
C-Major extinct	<input type="checkbox"/>	D-Attested extinct	<input type="checkbox"/>	E-Minor extinct	<input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic	<input checked="" type="checkbox"/>	Ideographic	<input type="checkbox"/>	G-Obscure or questionable usage symbols	<input type="checkbox"/>
4. Is a repertoire including character names provided? _____	No				

¹ Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

a. If YES, are the names in accordance with the “character naming guidelines”

in Annex L of P&P document? _____

b. Are the character shapes attached in a legible form suitable for review? _____

Yes

5. Fonts related:

a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?

TCA and related organizations

b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):

Already agreed with all TCA, Experts and related organizations

6. References:

a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

*Yes, see attached
PDF file*

b. Are published examples of use (such as samples from newspapers, magazines, or other sources)

of proposed characters attached?

Yes

7. Special encoding issues:

Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

No

. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database (<http://www.unicode.org/reports/tr44/>) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?

No

If YES explain

2. Has contact been made to members of the user community (for example: National Body,

user groups of the script or characters, other experts, etc.)?

No

If YES, with whom?

Experts of Small Seal, Education, Publishing Industry and seals.

If YES, available relevant documents:

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	No
Reference: <i>Evidence of use of these characters were reviewed by experts of Small Seal at the time of submission for Inclusion.</i>	
4. The context of use for the proposed characters (type of use; common or rare)	Common
Reference: <i>Widely used in community of Hanzi users for seals, books, decorations, etc..</i>	
5. Are the proposed characters in current use by the user community?	
If YES, where? Reference: Education, Publishing Industry and Stamp industry	
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	No
If YES, is a rationale provided?	
If YES, reference:	
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	Yes
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	No
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	No
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character?	No
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
11. Does the proposal include use of combining characters and/or use of composite sequences?	
If YES, is a rationale for such use provided?	No
If YES, reference:	
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	
If YES, reference:	

12. Does the proposal contain characters with any special properties such as

control function or similar semantics?

No

If YES, describe in detail (include attachment if necessary)

13. Does the proposal contain any Ideographic compatibility characters?

No

If YES, are the equivalent corresponding unified ideographic characters identified?

If YES, reference:

