Type:	**Working Group Document**
Title:	**Proposal to encode Four-Column Small Seal Script in UCS**
Source:	**TCA and China**
Reference:	**N4688, N5189, N5190, N5306, N5307**
Date:	**2025-06-06**

Summary

This document is rewritten based on n4688 and integrated with n5306 written by Michel to form a complete discussion. It explains the background significance, version selection, glyph correction principles, and encoding strategies for *Shuowen Jiezi* Small Seal Script into ISO/IEC 10646. The Small Seal Script encoding is primarily based on the Tenghuaxie version (THX), with some glyphs modified according to the "THX Seal glyph Correction Principles " (n5189). The code chart adopts a four-column layout, incorporating the Chenchangzhi version (CCZ), Xiaoxu version (QJZ), and Duan Zhu version (DYC). Due to glyph differences between versions, "Mapping Principles" (n5307) were established to ensure consistency and accuracy in encoding correspondence. Principles for correspondence with modern CJK characters are also proposed. This work establishes a systematic encoding framework for *Shuowen* Small Seal Script and provides a technical foundation for subsequent encoding of other ancient Chinese characters.

Introduction

The "Small Seal Script (小篆)" 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) to distinguish it 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), and it does not belong to the same system as modern Hanzi does. In addition, a large number of Small Seal Script characters cannot be mapped to modern standard Hanzi characters one to one; although some shapes can be mapped, the Small Seal Script characters have unique structures different from their counterparts in modern Hanzi. The Small Seal Script shall not be regarded as a typeface of modern Hanzi belonging 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 distinctive value to the study and research of ancient characters:

1. 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 researching ancient Hanzi characters.

2. The Small Seal Script characters in *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.

3. Having strictly collated and removed unnecessary glyphs, the Small Seal Script in *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*.

4. The 540 radicals summarized in the book *Shuowen Jiezi* cover all forms and meanings of Chinese characters, and are called structural radicals with theoretical values. Some later radicals 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 reference purpose, and are theoretically inadequate, the 540-radical is a more complete radical system.

Selection of Original Materials

The encoding of Small Seal Script is based on four different versions of *Shuowen Jiezi*, presented in a four-column layout. The repertory and ordering are primarily based on the Tenghuaxie (THX) version, with the Small Seal glyphs from the other three versions arranged according to the sequence of the THX version. Below is a brief introduction to the four versions of *Shuowen Jiezi* used:

1. THX: Tenghuaxie version (額勒布藤花榭本)

In Song Dynasty (宋朝, 960 AD to 1279 AD), Xu Xuan (徐鉉) was ordered by the Emperor to make a complete revision of the publication. He restored the contents and recovered its original glyphs. This version 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 the 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), Pingjinguan version (孫星衍平津館本) and Chenchangzhi version (陳昌治單行本, revised in accordance with Pingjinguan version). Both Tenghuaxie version and Pingjinguan version were made by following original Song Dynasty printed books, while Pingjinguan 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.

2. CCZ: Chenchangzhi version (陳昌治本)

The CCZ version was created by modifying the arrangement format of *Shuowen* based on Pingjinguan version, making *Shuowen* easier to read. As a result, it has gained recognition among scholars and has become an important reference for the study of *Shuowen Jiezi*.

3. QJZ: Xiaoxu Ben (小徐本)

The Xiaoxu Ben authored by Xu Kai (徐鍇) also has multiple versions, with the Qing Dynasty Daoguang (道光) period version by 祁雋藻刻本 widely recognized as the best. Therefore, this version serves as the basis for Xiaoxu Ben encoding.

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. For example, Tang Dynasty scholar Li Yangbing (李陽冰) altered the seal script characters and Xu Shen's explanations in *Shuowen*. Xu Kai of the Southern Tang (南唐) was the first person after Li Yangbing to revise and annotate *Shuowen Jiezi*, refuting Li Yangbing's alterations. Xu Kai's compilation of *Shuowen Xichuan* (說文繫傳) is called the "Xiaoxu Ben." (Xu Xuan was the elder brother and Xu Kai the younger brother, but the "Xiaoxu Ben" was completed during the Southern Tang period (AD 960-970), earlier than the "DaXu Ben" completed during the Northern Song (AD 986)).

4. DYC: Duan Zhu Ben (段注本)

Duan Yucai (段玉裁), a leading expert in *Shuowen Jiezi* research in Qing Dynasty, revised Daxu Ben and Xiaoxu Ben and then corrected some glyphs according to the examples of *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 *Shuowen Jiezi*, abbreviated as Duan Zhu (段注)). The Duan Zhu is an important basis for people now to learn *Shuowen Jiezi*. The different information between Daxu Ben and Duan Zhu was preserved based on Daxu Ben when the proposal was drafted.

Rules and Principles

1. Mapping Principles of Small Seal Glyphs in Different *Shuowen* Versions

The Small Seal Script encoding presents 4 versions in a 4- columns layout. The encoding follows the radical order and intra-radical character arrangement of the THX version as the primary basis. However, glyphs differ between versions, so horizontal encoding must consider whether glyphs from different versions can be unified. Unified glyphs can be encoded at the same code point; non-unified glyphs require separate encoding. For complete correspondence principles between versions, please refer to n5307. Summary as follows:

(1) General Principles:

A. **Version relationships:** The ordering, glyphs, and explanations of different versions of *Shuowen* are fundamentally consistent. The THX and CCZ version belong to the DaXu Ben, with largely identical content. The XiaoXu Ben and Duan Zhu differ from the DaXu Ben in some characters, where variations may occur.

B. **Coding sequence:** Arranged according to the order of appearance in the THX version. Those without correspondence are supplemented in sequential order.

C. **Written Form restrictions:** Only compares characters within the same writing form (e.g., Seal Script, Guwen, or Zhouwen); different forms are always processed as separate characters.

D. **Code Point Consistency:** Characters placed at the same code point must be consistent in both form and explanation; in case of conflict, character form takes

precedence.

(2) Unification Principle:

Characters meeting the following conditions are unified:

A. Taboo Glyphs

B. Those with incorrect seal script forms (glyphs inconsistent with explanations or lacking systematic component consistency)

C. Components are systematically modified in glyph

D. Variant characters with the same explanations but slightly different forms, like the four versions of "臧" in Small Seal script.
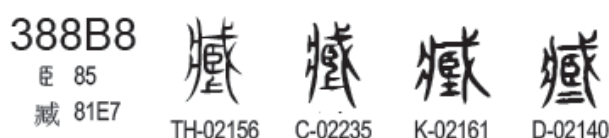
E. Phono-semantic compound structure variations (left-right or top-bottom positional interchange)

(3) Disunification Principle:

Characters meeting the following conditions are encoded separately:

A. Same pronunciation and meaning, but with differences in components , e.g., "藍, 蘫".

B. Compound characters with the same components and the same number of components, but with different structure.

C. single characters with the same pronunciation and meaning, but with differences in stroke count or stroke relationship (intersecting, connecting, or separate).

2. THX Seal glyph Correction Principles

Characters from THX version shall have shapes consistent with their original forms in the THX version. However, during organization of the Tenghuaxie Small Seal Script forms, scholars discovered character form problems. To ensure systematicity and consistency of the THX Small Seal Script character forms, corresponding

corrections have been made. Based on the nature of the character form problems, 6 correction principles can be summarized. For complete explanations, please refer to [n5189](). Summary as follows:

A. Taboo glyphs: Some glyphs seen as forbidden (such as those related to imperial names) appear in two ways: changed characters or omitting the strokes. For changed characters, they are restored by checking other versions; for characters with stroke omission, they are restored to their complete original forms.

B. Exact duplicate: *Shuowen Jiezi* contains characters with identical forms (and similar meanings). When such duplicate characters are all Zhengzhuan, only the character appearing earlier in sequence is encoded; when a group consists of Zhengzhuan and Chongwen with identical forms, only the Zhengzhuan is encoded, while the Chongwen is not. After comparison and organization, a total of 17 characters are not encoded due to identical forms.

C. The original broken glyphs: Due to rubbing or other conditions, some glyphs in the original book appear broken or incomplete. These glyphs are corrected based on the character's meaning and references from other versions, with the corrected glyphs used to avoid confusion.

D. The inconsistent component: Some glyphs contain inconsistent components. These inconsistencies are restored to original forms through comparison and etymological analysis according to the character's meaning.

E. Possible duplicate: Even glyphs that have only subtle differences are worth keeping and are encoded separately to maintain the integrity of the Small Seal script system.

F. Misprint of the original book: Some Small Seal script forms in *Shuowen Jiezi* are inconsistent with their explanations. By comparing with other versions, such as DaXu Ben, if confirmed as errors in the THX version, these incorrect forms are modified to align with their meanings.

3. **Correspondence to the modern CJK characters Principle**

A. For characters currently in Unicode encoding, use the clerical script form that

matches in both form and meaning as the corresponding modern character. For example, "示".

Serial No.	Serial No. in THX version	TTF	Corresp Modern Char.	SW Radical	Type
00017	00017	示	示	示	Zhengzhuan

B. For Zhengzhuan, when the corresponding modern character selected according to Rule 1 is not a commonly used standard character, the standard character must be supplemented. For example, "旁,旁".

Serial No.	Serial No. in THX version	TTF	Corresp Modern Char.	SW Radical	Type
00011	00011	旁	旁,旁	⊥	Zhengzhuan

C. For Guwen or Zhouwen, if there is no corresponding clerical script form exists in the current Unicode set, refer to the corresponding Zhengzhuan character instead.

Serial No.	Serial No. in THX version	TTF	Corresp Modern Char.	SW Radical	Type
02257	02178	殺	殺	殺	Zhengzhuan
02260	02181	殺	殺	殺	Variant（古文）
03463	03353	差	差	左	Zhengzhuan
03464	03354	差	差	左	Variant（籀文）

(End of document)

摘要

本文件依據 <u>n4688</u> 進行改寫,並參考 Michel 撰寫的 <u>n5306</u> 加以整合,成為完整的論述。本文說明《說文解字》小篆納入國際編碼的背景意義、版本選擇、字形校正原則與編碼排序策略。小篆編碼以藤花榭本(THX)為主,部分藤本字形依「<u>n5189</u>-校正原則」進行調整。碼表採多欄並列格式,搭配陳昌治本(CCZ)、小徐本(QJZ)與段注本(DYC)等版本。由於不同版本之間的字形略有差異,為確保編碼對應的一致性與正確性,制定「定序與定序原則」(<u>n5307</u>)。在對應今字方面,亦提出原則以供參照與應用。本項工作不僅為《說文》小篆建立系統性編碼架構,也為後續其他古漢字之編碼,奠定技術基礎與理論依據。

一、將《說文》小篆列入國際編碼的必要性

《說文解字》是中國漢代許慎編著的一部收集秦代小篆的專書。全書收集了小篆正篆 9353 字,異體字(《說文》稱"重文") 1163 字,重文中包括籀文(周代篆文)和古文(戰國時的六國文字)不與小篆相合的字樣。

秦代流行兩種正體字,一種是規範過的小篆,另一種是通常使用的隸書(文字學上稱作秦隸,也稱古隸,以與其後的漢隸有所區別)。有人認為小篆是一種書法字體,這是一種誤解。《說文》小篆是古文字發展中一個歷史階段的字符集,它與甲骨文、金文和戰國文字一樣,與現代楷書不屬於同一個體系,小篆字符有有大量的字形與現代楷書無法對應,能夠對應的字形在形體上也具有獨特的構造;所以,它不能作為現代漢字的一種字體而附屬於楷書,必須獨自佔有一個編碼區域。

《說文》小篆在漢字發展歷史上具有重要的作用,對古今文字的研究和整理更具有獨特的價值:

(一) 秦代小篆是古文字發展的最後階段,它囊括了自甲骨文以來先秦古文字的全部構造模式,與此前古文字的主要部件皆有聯繫,所以,它是古文字整理與考據必須參照的字符集。

(二) 《說文》小篆的字形與周代經典文獻密切相關,均可用經典文獻的意義來解釋構形理據,集中顯示了表意漢字的特點,也涵蓋了隸書以後今文字的字理。當前的漢字教學,都根據《說文》的講解來解讀漢字。

(三) 《說文》小篆經過嚴格整理，去除了冗餘的字形，形成了一個完整的構形系統，並將這些文字字符分屬在 540 個部首中，按小類編排了字符的順序。當前古文字整理，都以《說文》的部首和字序為序。

(四) 《說文》小篆歸納概括出的 540 部首，統領了漢字全部的形與義，被稱為具有理論價值的結構部首，之後的有些部首是在 540 部首基礎上簡化合併而成的，這種合併忽略了表意漢字形義統一的特點，僅僅適合查檢使用，在理論上是有缺欠的，540 部首才是更為完善的部首系統。

二、將《說文》小篆列入國際編碼的整理資源

小篆編碼共有 4 個《說文》版本來源，以 4 欄(four-columns)排列的形式呈現。編碼基準以藤花榭本為主，另外三個版本的小篆字形皆依藤本的順序進行排列。以下簡單介紹所採用的四種《說文》版本：

(一) 額勒布藤花榭本(THX 欄)

宋代徐鉉奉宮廷帝王之命，全面整理《說文》，恢復了它的本來面貌，得到了它的原初字形，徐鉉整理過的《說文解字》，稱作「大徐本」。這一版本為公認的規範版本，小篆編碼的字形與次序，首先當以此為據。

大徐本目前有三種比較常見的版本——額勒布藤花榭本、孫星衍平津館本，以及根據孫星衍平津館本修改的陳昌治單行本。藤花榭本與平津館本均依據北宋本刊刻，但平津館本實際上多據顧千里本小篆進行過刪改，不如藤花榭本更忠實于原本。根據目前情況，以藤花榭本作為底本，最為妥靠。

(二) 陳昌治本(CCZ 欄)

陳昌治本是根據孫星衍平津館本去修改《說文》的排列形式，讓《說文》變得較容易閱讀，因此也獲得研究者的重視，成為研究《說文》的重要依據。

(三) 小徐本(QJZ 欄)

徐鍇撰著的小徐本也有多種版本，公認以清道光年間祁嶲藻刻本為最好，因此以這個版本作為小徐本編碼的依據。

《說文》成書於東漢，經過漫長時間的流傳，真相有被湮滅的地方，如唐人李

陽冰，改動了《說文》的篆形與許慎的說解。南唐徐鍇是在李陽冰之後第一位對《說文解字》進行校訂與註釋，並駁斥李陽冰的改動；徐鍇整理的《說文繫傳》稱作「小徐本」。(徐鉉為兄，徐鍇為弟，但「小徐本」成書於南唐(小徐本成書年代約為 AD960~970)，早於成書於北宋的「大徐本」(大徐本成書年代為 AD986))。

(四) 段注本(DYC 欄)

清代《說文》學的權威段玉裁，校核大徐本與小徐本，並根據《說文解字》自身的體例，又改正了一部分字形，調整了一部分編排次序，作出了《說文解字注》(簡稱《段注》)，其中的成果，為現代人學習《說文》的重要依據。《段注》中與大徐本有區別的信息，需要在大徐本基礎上，加以保存。

三、整理條例 Rules and Principles

(一) 定字與定序原則

小篆編碼以 4 種版本、多欄並列方式呈現。排序以藤花榭本部首次序與部首內字符編排次序為主，但是各個版本之間的小篆字形有些差異，因此，在進行排序時，還要考量不同版本之間的字形是否認同；認同則可並列，若不認同，則需另外單獨編碼。有關不同版本間的對應原則，完整說明請見 n5307，摘要如下：

1.通則：

A. **版本關係**：不同版本《說文》的排序、字形及說解基本一致，藤花榭本與陳昌治本屬「大徐本」，內容大致一致。小徐本與段注本在部分字上與大徐本不同，可能出現差異。

B. **編碼順序**：依藤花榭本中出現的順序排列。無對應者，依序補入。

C. **形制限制**：僅比較同一書寫形制 (如篆文、古文、籀文)，不同形制視為不同字。

D. **同碼位一致性原則**：同一碼位字形與說解須一致，若有衝突則**以字形為主**。

2.認同原則。凡符合下述情形則予以認同，不再單獨編碼：

  A. 避諱字。

  B. 篆形錯誤（字形與說解不一致、構件不符系統性）。

  C. 系統性部件改形。

  D. 說解相同，字形僅有細微差異的異寫字，如「臧」各版小篆形。

388B8
臣 85
臧 81E7
TH-02156　C-02235　K-02161　D-02140

  E. 形聲字結構變化（左右、上下位置互換）

3.不認同原則。凡符合下述情形則分開編碼：

  A.音義相同但部件有差異（如"藍"和"灆"字）。

  B.合體字構件與數量相同但位置不同。

  C.獨體字音義相同但筆畫或筆畫關係不同（如相交、相離）。

(二)藤本字形校正原則：

取自藤花榭本之字，其字形須宗於該本原形，惟整理藤花榭本小篆字形的過程中，學者陸續發現小篆字形的問題。為使藤本小篆字形具有系統性與一致性，遂進行相應修正。根據字形問題的性質，可歸納出 6 種校正原則，完整說明請參見 n5189。摘要如下：

1.**避諱字(Taboo glyphs)：**

某些被視為禁忌的字形（如與帝王名字相關的字形）會被修改或省略筆畫。

2.**構形相同的字(Exact duplicate)：**

《說文解字》中存在構形相同(字義也相近)的字，當該組重複字都是正篆(Zhengzhuan)時，以排序在前的進行編碼；如是正篆(Zhengzhuan)與重文(Chongwen/Variant)的字形相同，則只編碼正篆(Zhengzhuan)，重文(Chongwen/Variant)不編碼。

因此，在經過整理比對之後，總共有 17 字因為構形相同的原故，而不予編碼。

3.殘斷字形(The original broken glyphs)：

由於拓印或其他原因，原書中的某些字形出現殘缺。這些字形會根據字義和其他版本的參考進行更正，並使用更正後的字形來避免混淆。

4.不一致之部件(The inconsistent component)：

某些字形出現不一致的部件，這些不一致會通過比對和字理恢復到原始字形。

5.近似字處理(Handling of possible duplicate)：

即使字形差異非常微小，但仍值得保留這些字形，並將其單獨編碼。這是為了保持小篆字形系統的完整性。

6. 原書誤刻(Etymological errors in the original book)：

《說文解字》中有些小篆字形的字形與解釋不一致，經與其他版本比對後，如果確認是藤花榭本的誤刻，就會這些字形將進行修正，以符合字義。

(三)對應今字原則：

1.以目前 Unicode 編碼字中，與之形、義相符的隸書形為對應今字。如「示」。

Serial No.	Serial No. in THX version	TTF	Corresp Modern Char.	SW Radical	Type
00017	00017	示	示	示	Zhengzhuan

2.若是正篆，當按規則 1 選擇的對應今不是常見的通用字形時，需補上通用字形。如「旁,旁」。

Serial No.	Serial No. in THX version	TTF	Corresp Modern Char.	SW Radical	Type
00011	00011	旁	旁,旁	上	Zhengzhuan

註：「旁」-通用字形；「旁」-形、義相符的隸書形

3. 對於古文/籀文，若在目前 Unicode 編碼字中沒有與之對應的隸定形體，先依正篆對應字。

Serial No.	Serial No. in Tenghuaxie version	TTF	Corresp Modern Char.	SW Radical	Type
02257	02178		殺	殺	Zhengzhuan
02260	02181		殺	殺	Variant（古文）
03463	03353		差	左	Zhengzhuan
03464	03354		差	左	Variant（籀文）

(End of document)