

Doc Type: Working Group Document
Title: Shuowen Seal Encoding Report
Source: TCA
Status: Proposal
Action: For consideration by JTC1/SC2/WG2 and UTC
Reference N4853、N5034、N5089、N5108
Date: 2019-09-27

After Shuowen Seal was proposed from [N4634](#) (2014-09-30), there are about 20 proposals. This document summarizes the consensus for Shuowen Seal encoding according to the 4 reports:

- (1) Adhoc meeting on 2017/08/28-29 (Taipei, Taiwan) with the document [N4853](#),
- (2) WG2#67 meeting(London, UK; 2018/06/18-22) with the document [N5034](#),
- (3) the Informal meeting on 2019/6/1 (Beijing, China) with the document [N5089](#),
- (4) WG2#68 meeting (Redmond, US; 2019/06/17-21) with the document [N5108](#);

The consensus for Shuowen Seal encoding includes 9 issues.

《說文》小篆編碼從 N4634(2014-09-30)正式提案迄今，約有 20 篇提案文件。這份文件是將目前的編碼共識加以整理，以作確認，主要依據的決議報告有四份：

- (1) 2017 年 8 月 28-29 日的 Adhoc meeting (Taipei, Taiwan)，決議報告為 N4853；
- (2) WG2#67(2018/06/18-22, London)決議，N5043；
- (3) 2019 年 6 月 1 日的臨時會議(Beijing, China)，決議報告為 N5089；
- (4) WG2#68(2019/06/17-21, Redmond)決議，N5108；

有關小篆編碼的共識整理成 9 個議題，說明如下：

1. Handling of Shuowen Seal encoding (《說文》小篆編碼方式).....	2
2. The version of Shuowen Seal encoding (《說文》小篆編碼的版本依據)	2
3. Handling of naming 《說文》小篆碼區命名.....	3
4. Handling of taboo glyphs (避諱字)	3
5. Handling of Chongwen/ Exact duplicate (重文)	5
6. Handling of possible duplicate (近似字).....	7
7. Handling of etymological errors in the original book(原書誤刻)	8
8. Handling of the original broken or inconsistent component (殘斷或不一致之部件)...	9
9. Annex S.....	13

1. [Handling of Shuowen Seal encoding \(《說文》小篆編碼方式\)](#)

Refer to WG2 [N4853](#). And has been accepted in the WG2 66th meeting (Hohhot, China; 2017-09-25/29), See WG2 [N4874](#) Recommendation M66.20 (Shuowen Small Seal script).

There is unanimous agreement that Shuowen Seal characters should be encoded as **a separate script, in a new block**, independent of CJKUI (and not by means of a VS scheme taking CJKUI as Base), beginning with “U+4XXXX” (per the Road Map, in the TIP=Tertiary Ideographic Plane).

2. [The version of Shuowen Seal encoding \(《說文》小篆編碼的版本依據\)](#)

2.1 In two steps

For Shuowen Seal encoding, our consensus is divided into two steps. First, encoding Shuowen Seal as independent characters based mainly on 藤花樹本 version, and to expand the Standard Code Chart by adding non-unifiable glyphs from the other editions (such as 陳昌治本, 段注本, etc.) in steps.

對於《說文》小篆編碼，大家的共識為分兩階段進行；先以藤花樹本為基礎，編碼後，再與其他版本(如陳昌治本、段注本等)進行比較，不同於藤花樹本的字形則予以擴編。

2.2 藤花樹本 serve as “baseline”

Also the resolution reached at the last Adhoc meeting (aka WG2 [N4853](#) Q3) :

There is agreement that 藤花樹本, preserved in 東京專門學校 (which is now renamed as 早稻田大學), **may serve as “baseline”**, for the determination of the repertory (the number of rows / code points), and for production of **a single-column chart in 藤花樹本 glyph style**.

In WG2 68th meeting (Redmond, US; 2019-06-17/21), we have the consensus on the policy of encoding: According to *Shuowen* preface, Xu Shen(許慎) wrote as “今敘篆文，合以古籀.” The Guwen and Zhouwen collected in *Shuowen Jiezi* are classified as Shuowen Guwen (《說文》古文) or Shuowen Zhouwen (《說文》籀文). **Therefore, this policy is only for Shuowen Seal, Shuowen Guwen and Shuowen Zhouwen.** On the other hand, the characters in Hanjian(汗簡) or Three Script Stone(三體石經) are classified as Chuanchao Guwen (傳抄古文). These characters are out of encoding range. See WG2 [N5089](#) & [N5108](#).

3. [Handling of naming 《說文》小篆碼區命名](#)

According to the discussion in the WG2#67 meeting, and the report of N5034, the script block name was decided to use “Small Seal”. See as follows:

At the last, the script block name is discussed; the candidates are Shuowen Small Seal, Small Seal or Seal. The majority of WG2 experts supported Small Seal. Suzuki concerned that Guwen and Zhouwen are not Small Seal, thus “Seal” is the best. Selena responded Guwen and Zhouwen are Small Seal. In this meeting, **it was decided to use the name “Small Seal”**, however, it could be revised when non-Shuowen sourced characters are added in future.



依據 wg2#67 的討論，以及 N5034 的紀錄，決定《說文》小篆碼區名稱為「Small Seal」。N5034 的原文如上。

4. [Handling of taboo glyphs \(避諱字\)](#)



According to expert verification, the taboos of Tenghuaxie version are presented in two ways of changing the character or omitting the strokes, a total of 11 taboo characters (see [N4973](#) & [N5089](#)). See below.

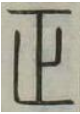


There are two ways to deal with it:

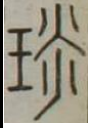

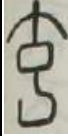


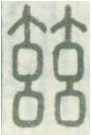


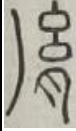







A. if it is confirmed “changing the character”, please refer to the other







version to restore it, such as “正”  →  ;

B. If it is confirmed “omitting the strokes”, use the original glyph as the

prefix, and treat other characters by IVS/IVD., such as “玄”  → .

No.	Vol.	Serial No.	Serial No. in Tenghuaxie version	Original Glyph	TTF Font	Refer to 陳本 to restore the glyph	Corresp Modern Char.	SW Radical	SW Radical Number	Type
1.	卷1	23	00023				禛	示	3	Zhengzhuan

No.	Vol.	Serial No.	Serial No. in Tenghuaxie version	Original Glyph	TTF Font	Refer to 陳本 to restore the glyph	Corresp Modern Char.	SW Radical	SW Radical Number	Type
2.	卷1	144	00140		琰		琰	玉	6	Zhengzhuan
3.	卷4	475	02773		玄		玄	玄	126	Zhengzhuan
4.	卷4	477	02775		兹		兹	玄	126	Zhengzhuan
5.	卷4	478	X096		𠂔			玄	126	Newly added character
6.	卷4	634	02931		胤		胤	肉	135	Zhengzhuan
7.	卷7	90	X155		曆			日	231	Newly added character
8.	卷7	209	04828		鎡		鎡	鼎	250	Variant
9.	卷8	455	x202		𠂔			衣	300	Newly added character

No.	Vol.	Serial No.	Serial No. in Tenghuaxie version	Original Glyph	TTF Font	Refer to 陳本 to restore the glyph	Corresp Modern Char.	SW Radical	SW Radical Number	Type
10.	卷9	38	06178				顚	頁	324	Zhengzhuan
11.	卷11	173	07771				洙	水	410	Zhengzhuan






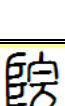
5. Handling of Chongwen/ Exact duplicate (重文)

These Chongwen sets are merely different in line shape and length, and they have no influence on glyph structure. It has been confirmed **only one code point will be encoded**. (see [N4973](#) & [N5089](#))

After verification by experts, there are 15 sets of Chongwen. Please see below for the full contents.

No.	Serial No. in Tenghuaxie version	SW Radical	Corresp Modern Char	TTF Font	pronunciation	meaning	Note
1	00939	口	右		you4	助也。从口从又。	encode
	02078	又	右		you4	手口相助也。从又从口。	
2	01005	口	否		fou3/pi3	不也。从口从不。	encode
	08352	不	否		fou3/pi3	不也。从口从不，不亦聲。	
3	00888	口	吹		chui1	嘘也。从口从欠。	encode
	06060	欠	吹		chui1/chui1	出气也。从欠从口。	

No.	Serial No. in Tenghuaxie version	SW Radical	Corresp Modern Char	TTF Font	pronunciation	meaning	Note
4	00980	口	吁		xu1	驚也。从口于聲。	encode
	03407	亏	吁		xu1	驚語也。从口从 亏，亏亦聲。	
5	02780	放	敖		ao2	出游也。从出从 放。	encode
	04307	出	敖		ao2	游也。从出从放。	
6	03436	豈	愷		kai3	康也。从心、豈， 豈亦聲。	encode
	07327	心	愷		kai3	樂也。从心豈聲。	
7	07729	水	湫		tuo1/tuo4	河津也。在西河 西。从水叕聲。	encode
	00881	口	湫		tuo4	口液也。从口垂 聲。唾或从水。	
8	07347	心	愬		zhe2	敬也。从心折聲。	encode
	00899	口	愬		zhe2	知也。从口折聲。 哲或从心	
9	06118	欠	𢇛		you3	愁兒。从欠幼聲。	encode
	01029	口	𢇛		you1	鹿鳴聲也。从口幼 聲。呦或从欠	
10	10374	車	輟		chuo4	車小缺復合者。从 車𠂔聲。	encode
	05346	网	輟		zhuo2	捕鳥覆車也。从网 𠂔聲。𠂔或从車	
11	03072	刀	剗		duo2	判也。从刀度聲。	encode
	02259	支	剗		du4	閉也。从支度聲。 讀若杜。𠂔或从	

No.	Serial No. in Tenghuaxie version	SW Radical	Corresp Modern Char	TTF Font	pronunciation	meaning	Note
						刀。	
12	03915	木	梃		ni3	木也。實如梨。从木尼聲。	encode
	04191	木	梃		chi4/ni3	簍柄也。从木尸聲。尿或从木尼聲。	
13	08507	手	擗		gong3	擗也。从手巩聲。	encode
	02063	𠂇	擗		gong3	衰也。从𠂇工聲。𠂇或加手。	
14	09582	虫	螯		mao2	蟹螯也。从虫𠂇聲。	encode
	09707	𧈧	螯		mou2/mao2	蟲食艸根者。从蟲，象其形。吏抵冒取民財則生。蠹或从𧈧。	
15	10500	𠂇	院		yuan4	堅也。从𠂇完聲。	encode
	05054	阝	院		huan2/yuan4	周垣也。从阝奂聲。窳或从𠂇。又爰眷切。	

6. [Handling of possible duplicate \(近似字\)](#)

In principle, even if glyphs that have only subtle differences are worth keeping, and they would be encoded individually. See WG2 [N4973](#).

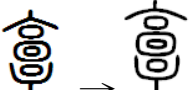
But in the [N5089 Q3](#), two sets of very similar glyphs were adjusted. Due to Xu Shen clearly stated that Guwen and Zhouwen should not be resemble. To improve the glyphic difference, **the representative glyphs would be updated by using other versions**, as explained in the following table:





























A. Tenghuaxie version *Shuowen Zhouwen* is changed to Duan

version (段注本) glyph :

王	孫	陳	N4688	TCA	毛4	毛5	華東	北師	逢甲	段	述	汪	祁
													
06957	06957	06957	06956	06957	06955	06960	07906	13585	26861	06686	06714	06477	06712
													
06958	06958	06958	06957	06958	06956	06961	07907	13586	3701	06687	06715	06478	06713
													
06959	06959	06959	06958	06959	06957	06962	07908	13649	3700	06688	06716	06479	06714

B. Tenghuaxie version *Shuowen Zhouwen* is changed to 北師大小

篆 font :

													
03832	03832	03832	03832	03832	03830	03833	0480	13768	12916	03698	03723	03709	03723
													
10208	10208	10208	10202	10208	10205	10213	11155	14254	82916	09836	09824	09508	09857

7. Handling of etymological errors in the original book(原書誤刻)

After verification by experts, it was found that the original book made etymological errors in five characters. For most characters, *Shuowen* mentions each component. In five problematic characters, the components in the explanation and the representative glyphs show inconsistencies. By the comparisons with other DaXu versions of *Shuowen*, they are confirmed to be the errors in Tenghuaxie version. For these 5 characters, the corrected representative glyphs would be used in the proposal, instead of original wrong glyphs. See WG2 [N5089 Q4](#).

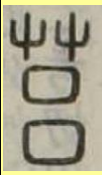






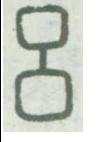



No.	Vol.	Serial No.	Serial No. in Tenghua xie version	Original glyph	TTF Font	Corresp Modern Char.	SW Radical	SW Radical Number	Type	Content
1	卷 1	650	00632			莢	艸	12	Zheng-zhuan	從艸次聲
2	卷 10	392	07082			焯	火	382	Zheng-zhuan	原圖疑誤刻，从火皂聲，字樣依陳本修正
3	卷 11	703	08271			鮎	魚	424	Zheng-zhuan	原圖疑誤刻，从魚占聲，字樣依段注本修正
4	卷 13	509	09700			蠹	虫	472	Zheng-zhuan	陳本跟段注本皆出頭
5	卷 8	319	05767			臥	卧	297	Zheng-zhuan	原圖疑誤刻，从人、臣，取其伏也。

8. [Handling of the original broken or inconsistent component \(殘斷或不一致之部件\)](#)

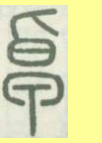


After verification by experts, it was found that there are **11 characters** contains broken or inconsistently designed glyphs in the original book due to rubbing or other conditions; the finding was mainly based on the same component glyphs of the same version. Therefore, correction was processed according to the meaning of the character and the reference of other versions. Since showing broken or inconsistently designed glyphs as the representative glyphs would be confusing for the implementers, the corrected representative glyphs would be used in the proposal. Also no participants supported the separate encoding of the original but broken glyph or inconsistently designed glyphs at the different codepoints. To use such glyphs for back compatibility, teaching, comparison etc,

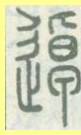



















the participants recommended to use IVD/IVS solution. See WG2 [N5089 Q5](#).

I. 「呂」

	No.	Vol.	Serial No.	Serial No. in Tenghuaxie version	Original glyph	TTF Font	Corresp Modern Char.	SW Radical	SW Radical Number	Type	Content
Need to restore	1	卷1	321	00303			莒	艸	12	Zheng-zhuan	依字理恢復原形 《說文解字·呂部》:「  ，脊骨也。象形。」
Component of “呂”	A.	卷5上	60	03235			筭			Zheng-zhuan	
	B.	卷6上	247	04071			𣎵			Zheng-zhuan	
	C.	卷7上	517	05124			呂			Zheng-zhuan	
	D.	卷12上	45	08391			閭			Zheng-zhuan	

II. 「卓」

	No.	Vol.	Serial No.	Serial No. in Tenghuaxie version	Original glyph	TTF Font	Corresp Modern Char.	SW Radical	SW Radical Number	Type	Note
Need to restore	2	卷8	290	05738			卓			Zheng-zhuan	依字理恢復原形 早、匕為卓。而早，从日在甲上。甲，作  形。

	3	卷2	563	01302			卓			Zheng-zhuan	See “卓”
Component of “卓”	1.	卷8	72	05538			倬			Zheng-zhuan	
	2.	卷10	396	07086			焯			Zheng-zhuan	
	3.	卷10	885	07567			悼			Zheng-zhuan	
	4.	卷7	725	05332			罩			Zheng-zhuan	
	5.	卷12	688	09015			焯			Zheng-zhuan	
	6.	卷12	277	08617			掉			Zheng-zhuan	
	7.	卷11	254	07852			淖			Zheng-zhuan	
	8.	卷2	361	01101			趙			Zheng-zhuan	
	9.	卷11	751	08321			鯈			Zheng-zhuan	

III. Others

No.	Vol.	Serial No.	Serial No. in Tenghuaxie version	Original glyph	TTF Font	Corresp Modern Char.	SW Radical	SW Radical Number	Type	Note	Note
-----	------	------------	----------------------------------	----------------	----------	----------------------	------------	-------------------	------	------	------

No.	Vol.	Serial No.	Serial No. in Tenghuaxie version	Original glyph	TTF Font	Corresp Modern Char.	SW Radical	SW Radical Number	Type	Note	Note
4	卷5	51	03226				籊	竹	143	Zheng-zhuan	原圖拓印不全，依陳本正形
5	卷5	283	03452				𪔐	𪔐	166	Zheng-zhuan	右上從宁，原圖不應斷筆
6	卷13	522	09713				蜚			Variant	原字形當从非，中間不應斷
7	卷13	746	x380				坊	土	480	Newly added character	依形从土方聲，不宜斷開
8	卷13	809	09983				勸	力	488	Zheng-zhuan	虎形
9	卷14	251	10273				𪔐	斗	496	Zheng-zhuan	少一點
10	卷14	253	10275				𪔐	斗	496	Zheng-zhuan	殘筆
11	卷7	215	04834				𪔐	𪔐	252	Zheng-zhuan	圈起處應該連起來。

9. [Annex S](#)

According to the report of the N4853 Q1, it written as:

As a separate script, CJKUI guidelines (such as Annex S) will not apply, but new rules will be developed and documented in the process of determining the repertory.

If we think it is necessary in the future, let us discuss it again.

根據N4853決議，不要有Annex S。未來如果認為有其必要，再作討論。

(End of Document)