1. Introduction

Etymologists divide the history of Chinese scripts development into contemporary stage and ancient stage by the delimiter of the Liban (the seal script change).   Nowadays, the Internet provides a newly cyber space.   Facing the transitions of the Chinese languages and characters, there exist platforms for contemporary scripts appeared after Liban.   Traditional-Simplified and form-variant character exchange mechanisms are available globally across country boarders. However, in order to set up a platform for ancient script digital processing, an encoding system for ancient script message exchange is needed for information process.   A working group for ancient script encoding was agreed to be formed according to the conclusion of the 2003 IRG meeting held in Chengdu. In that meeting, the consensus of proper attention to both ancient and contemporary scripts was reached.   Also, the scope of old Hanzi is bounded to Pre-Qin scripts appeared before the Liban.   The conclusion was approved by WG2 meeting held in Canada, June 2004.   TCA thus, gathering computer experts and ancient script scholars, formed a working group for ancient script encoding.   The working group aggressively starts to discuss ancient script encoding rules and plan the flow of pre-work.   The progress will be reported to the incoming IRG meeting in Cheju, Korea, asking opinions for the further work.

2. Encoding Reasons

2.1 Old Hanzi should be classified based on the principles of the time they were used, the locations they were collected, the materials they were on, their form-variances, and properties of form, pronunciation, and meaning.   This establishes the baselines for old Hanzi encoding classification.

2.1.1 Period: Shang Dynasty, Western Zhou Dynasty, Spring and Autumn Period, Warring Kingdoms, and Qin Dynasty.

2.1.2 Materials: Jia-gu-wen (inscription on animal bones and tortoise shells) (divided into "the Shang Dynasty Jia-gu-wen " and "the Western Zhou Dynasty Jia-gu-wen), Jinwen (bronze inscription) (including Zhou-wen), jade and stone writings (jade carving, stone carving and so on, including other carving), bamboo and silk writings (bamboo books(Jan-tu) and silk books.   Qin-li is part of "Jan-tu"), small Juahn (majority are "Shuowen Jiezi" small Juahns, covering "Shuowen Sinfu" small Juahns), and miscellanea (including ancient pottery writings, seal writings, currency writings, lacquer ware writings, etc.).

2.1.3 Locations: In Warring Kingdoms period, there were five character affiliations, namely Qin (秦), Chu(楚), Jin(晉), Qi(齊), and Yan(燕).

2.1.4 Writing-variances: Variant constructs for a character need to be encoded separately.   In the case of variant writing, the representative writing should be chosen.   Other writing-variances should be listed aside with the representative. Characters with variant constructs may be listed as notes.   A construct-variance replaces a component in a character with a different component or a component with different writing.   The writing-variance involves some stroke(s) in a character having minor different drawing(s), however the basic structure of the character is retained.   The evolution of old Hanzi took a long time.   The decision involves the structure differences of basic strokes.   For example, if a dot in a character became a horizontal bar, the character is deemed as a construct- variance.   However, characters with different sizes of dot of different lengths of horizontal bar will make them writing-variances.

2.2 Ancient script encodings are divided into six categories and twenty two blocks based on classification principles stated above.

Jiaguwen (甲骨文)	1 Shang Jiaguwen (商甲骨文)	2 Western Zhou Jiaguwen (西周甲骨文)
Jinwen (金文)	3 Shang Jinwen (商金文)	4 Western Zhou Jinwen (西周金文)
	5 Spring and Autumn Period Jinwen (春秋金文)	6 Warring Kingdoms Chu Jinwen (戰國楚系金文)
	7 Warring Kingdoms Jin Jinwen (戰國晉系金文)	8 Warring Kingdoms Qi Jinwen (戰國齊系金文)
	9 Warring Kingdoms Yan Jinwen (戰國燕系金文)	10 Warring Kingdoms Qin Jinwen (including Zhouwen) (戰國秦系金文(含籀文))

Jade and stone writings (玉石文字)	11 Shang Jade and stone writings (商玉石文字)	12 Jade and stone writings (西周玉石文字)
	13 Spring and Autumn Period Jade and stone writings (春秋玉石文字)	14 Warring Kingdoms Chu Jade and stone writings (戰國楚系玉石文字)
	15 Warring Kingdoms Jin Jade and stone writings (戰國晉系玉石文字)	16 Warring Kingdoms Qi Jade and stone writings (戰國齊系玉石文字)
	17 Warring Kingdoms Yan Jade and stone writings (戰國燕系玉石文字)	18 Warring Kingdoms Qin Jade and stone writings (戰國秦系玉石文字)
Bamboo and silk writings (簡帛文字)	19 Warring Kingdoms Chu bamboo and silk writings (戰國楚系簡帛文字)	20 Warring Kingdoms Qin bamboo and silk writings (戰國秦系簡帛文字)
Small Juahn (小篆)	21 Small Juahn (小篆)	
Miscellanea	22 ancient pottery writings, seal writings, currency writings, lacquer ware writings, etc. (古陶文、璽印文字、貨幣文字、漆器文字)	

3. Encoding Process

3.1 Rules of character selection

3.1.1 Representative characters: Representative characters were chosen by experts. Rules for collecting representative characters are as follow: well-recognized characters are collected; uncertain or arguable characters are left out. Representative characters are manually drawn careful with aesthetic and uniform styles.　Original scripts are scanned into graphic files with clear resolutions.　Please see the appendix: Ancient Jinwen script encodings table.

3.1.2 Decision of construct-variances: Construct-variances are deemed as different ancient characters, thus they should be encoded separately.　In the case of variant writing, the representative writing should be chosen.　Other writing-variances should be listed aside with the representative.　Characters with variant constructs may be listed as notes.　Only the most five popular Construct-variances will be collected for each character.

3.1.3 Writing-variances should be affirmed by experts.　Rules of their recognition should be logged character by character.

3.2 Encoding procedures

3.2.1 Fill in a row in the attribute table shown below for each character collected first. This table will be used as the foundation for encodings in each block.　(Please see examples in the appendix.)

| Running No.	Rep. Script	Original Shape	Source	Carrier	Period	Area	Material	Radical	Pronunciation	Glyph Determ.	Corresd. Modern Char	Notes

Data attributes:

3.2.1.1 Running number: used to identify individual script.

3.2.1.2 Representative script: Provides a reference script for comparison based on emulating drawing from the original material.

3.2.1.3 Original shape: Original scripts on archaeological finds or their rubbings.

3.2.1.4 Source (category): Records literature, page number, sequence number or index of the source of the script.　(The sources are presented with abbreviations. The abbreviation table is given.)

3.2.1.5 Carrier：Name of the archaeological find or literature carried the script.

3.2.1.6 Period：The period the carrier of the script was used or created.　There were five periods: Shang Dynasty, Western Zhou Dynasty, Spring and Autumn Period, Warring Kingdoms, and Qin Dynasty.

3.2.1.7 Area: The place the carrier was used or created.

3.2.1.8 Material: The material of the carrier of the script.

3.2.1.9 Radical: coded with three digits, corresponding to 540 radicals in "Shuowen Jiezi."

3.2.1.10 Pronunciation：mark the pronunciation of the script with phonetic symbols.

3.2.1.11 Glyph determination: Converts a script into a corresponding Kaishu glyph according to the rules of glyph determination.

3.2.1.12 Corresponding modern characters: Lists one or more possible corresponding modern character(s) of the script.

3.2.1.13 Notes: Related comment.

3.2.2 Organization and encoding sequence is as follow: Jiaguwen, Jinwen, jade and stone writings, jade and stone writings, bamboo and silk writings, small Juahn, and Miscellanea.

4. Ending words

4.1 Experiencing ancient script encoding workflow construction, the working group clarified attributes of shapes of ancient Chinese scripts, distinguished construct-variances from writing-variances, and confirmed recognition rules of shape variances.　The working group encoded old Hanzi in order based on attribute definitions.　The working group constructed an ancient script encoding system to facilitate digital character exchange in Chinese character cultural circle.

4.2 The task of ancient script encoding is arduous; the questions related to it are complicated.　If the task can be accomplished stage by stage, ancient literatures and archaeological found literatures will be able to be retrieved and printed.　This will accelerate the speed of recording and organizing cultural heritages, and will accelerate publication and propagation of research results.

Appendix:

Shang Jinwen Script Encoding Table

Running No,	Rep. script	Original shape	Carrier	Source	Period	Area	Material	Radical	Pronunciation	Glyph determ.	Corresd. modern char.	Notes
001			狽元作父戊卣[1]	《集成》5278.2-6	Shang		bronze	元	yuan	元	元	或釋為兀
002			△天父癸簋 [2]	《集成》3340	Shang		bronze	天	tian	天	天	

Western Zhou Jinwen Script Encoding Table

Running number	Rep. script	Original shape	Carrier	Source	Period	Area	Material	Radical	Pronunciation	Glyph determ.	Corresd. modern char.	Notes
001			我方鼎	《集成》2763.1-4	W. Zhou		bronze	一	yi	一	一	
002			智鼎	《集成》2838-6	W. Zhou		bronze	元	yuan	元	元	
003			沈兒鐘	《集成》203.2-9	W. Zhou		bronze	元	yuan	元	元	
004			天作从尊	《集成》5688	W. Zhou		bronze	天	tian	天	天	
005			頌壺	《集成》9731	W. Zhou		bronze	天	tian	天	天	

---

[1] 器名舊作〈兀作父戊卣〉。
[2] 字形見於〈△天父癸簋〉。

Running No.	Rep. script	Original shape	Carrier	Source	Period	Area	Material	Radical	Pronunciation	Glyph determ.	Corresd. modern char.	Notes
001	一		秦公簋	《集成》4315.3-9	Spring & Autumn	Qin	bronze	一	yi	一	一	
002			朐簋		Spring & Autumn		bronze	元	yuan	元	元	
003			秦公簋	《集成》4315.3-9	Spring & Autumn		bronze	元	yuan	元	元	
004			少虞劍	《集成》11696	Spring & Autumn		bronze	元	yuan	元	元	
005			虞公劍	《集成》11663	Spring & Autumn		bronze	元	yuan	元	元	
006			吳季子之子劍	《集成》11640	Spring & Autumn		bronze	元	yuan	元	元	
007			洹子孟姜壺	《集成》9730	Spring & Autumn		bronze	天	tian	天	天	
008				《集成》211	Spring & Autumn		bronze	天	tian	天	天	

The table title is: **Spring and Autumn Period Jinwen Script Encoding Table**

---

³ 器名舊稱〈吉日壬午劍〉。
⁴ 器名舊稱〈粦公劍〉。

Warring Kingdoms Jin Jinwen Script Encoding Table

Running No.	Rep. script	Original shape	Carrier	Source	Period	Area	Material	Radical	Pronunciation	Glyph determ.	Corresd. modern char.	Notes
001	一		十一年皋落戈	《考古》1991年第5期	Warring Kingdoms	Jin	bronze	一	yi	一	一	
002			元年鄭令矛	《集成》11552	Warring Kingdoms	Jin	bronze	元	yuan	元	元	
003			中山王△鼎	《集成》211	Warring Kingdoms	Jin	bronze	天	tian	天	天	
004			中山侯恣鉞	《集成》11758	Warring Kingdoms	Jin	bronze	天	tian	天	天	

## Warring Kingdoms Yan Jinwen Script Encoding Table

Running No.	Rep. script	Original shape	Carrier	Source	Period	Area	Material	Radical	Pronunciation	Glyph determ.	Corresd. modern char.	Notes
001	一		重金扁壺	《集成》9617-8	Warring Kingdoms	Yan	bronze	一	yi	一	一	

Warring Kingdoms Chu Jinwen Script Encoding Table

Running No.	Rep. script	Original shape	Carrier	Source	Period	Area	Material	Radical	Pronunciation	Glyph determ.	Corresd. modern char.	Notes
001			鄂君啓舟節	《集成》12113	Warring Kingdoms	Chu	bronze	能	yi	△	一能	字從能翼省聲

Warring Kingdoms Qin Jinwen Script Encoding Table

Running No.	Rep. script	Original shape	Carrier	Source	Period	Area	Material	Radical	Pronunciation	Glyph determ.	Corresd. modern char.	Notes
001	一 5		卅一年相邦冉戈	《集成》12342	Qin	Qin	bronze	一	yi	一	一	

Warring Kingdoms Qi Jinwen Script Encoding Table

Running No.	Rep. script	Original shape	Carrier	Source	Period	Area	Material	Radical	Pronunciation	Glyph determ.	Corresd. modern char.	Notes
001	亓		陳逆簠	《集成》4630	Warring Kingdoms	Jin	bronze	元	yuan	元		

5 原形參王輝編著《秦銅器銘文編年集釋》圖四十七，頁 57。舊稱〈廿一年相邦冉戈 〉