Universal Multiple-Octet Coded Character Set UCS	

The Old Hanzi principles and references (version 2) extracted and compiled from the following documents:

IRG N1135R, IRG N1182, IRGN1215, IRG N1267.

1. Format of submission

ID	Rep. Script/ Glyph	Original Shape/ Glyph	Source	Period/ Epoch	Area/ site	Material	SW Radical	SW Radical Number	Glyph Determ.	* Corresp. Modern Char (UCS Code)	* Unifiable Shapes	* Notes
1												
2												

2. Definitions

2.1 ID: It is the unique id that consists of one or two letters member id (G, T, K, KP, J, V, S, H, M) followed by four digit sequential numbers assigned by submitters.

Example: T0001 is one IRG global unique ID assigned to an Old Hanzi submitted by TCA.

2.2 Rep. Script/Glyph: The truthful trace from 'Original Shape/ Glyph'

[代表字的定義：忠實摹寫原形的字形。]

2.3 Original glyph: The glyph selected according to the principles of Old Hanzi selection (in the item 4) in this document.

[原形的定義：根據選字原則(本文件第 3 章)選定的字形。]

2.4 SW Radical: ShuoWen Radical

2.5 SW Radical number : 1 – 540

2.6 The "Glyph determination (Liding)": It is a transcription of ancient script forms originally into 'clerical-style forms' (Li style), and today into Kai style with preservation of the shapes of the former. For this "Glyph determination (Liding)", the extended Kai style forms will be used without limitation of known Kai strokes.

[隸定字的定義 :將古漢字的原形用楷書或隸書的筆畫加以轉寫成的字形。]

3. Rules

3.1 The 'Rep. Script/Glyph' should be of standardized size and dimension.

3.2 The 'Rep. Script/Glyph' shall be in EPS format (resolution 1024×1024).

3.3 The format of bitmap images for rep. glyphs, original shapes and 540 SW Radicals are specified as follows:

128x128，Black and white bitmap

3.4 The last three columns are 'optional' field and they are indicated with an asterisk "*", the other fields are mandatory.

3.5 If the field "Corresp. Modern Char.(UCS Code)" is blank, then the "Notes field" must be filled with justifications to indicate the glyph is well-understood, for example, the meaning of the 'Rep. Script / Glyph'.
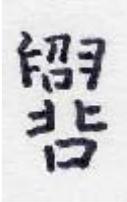
3.6 For the field 'Corresp. Modern Char.(UCS Code)' there may be no corresponding Simplified character only corresponding Traditional characters.

3.7 Glyph determination (Liding)

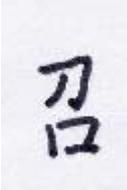(a) Retain the number of the components of the original scripts. Examples are shown below.

[保持原形構字部件的數量。]

Original Script/Glyph	Rep. Script/Glyph	Glyph determination (Liding)
		3 components
		8 components

		7 components

(b) Retain the relative positions and faces of the components of the original scripts. Examples are shown below.

[保持原形構字部件的相對位置及方向。]

Original Script/Glyph	Rep. Script/Glyph	Glyph determination (Liding)
An example of two inscriptions with different face of the same component. the top component in the first inscription is a mirrored image of '刀'.		
An example of two inscriptions which consists of components with different relative position.		

		靳
		辠

(c) In the transcription of Oracle Bone forms into Kai style, known Kai strokes must be used when there are equivalents in Kai style. Examples are shown below.

[甲骨文的構字部件如在楷書中已有對應，即用楷書部件形體轉寫。]

Original Script/Glyph	Rep. Script/Glyph	Glyph determination (Liding)
		召
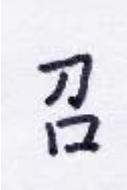
(d) The glyph should be in PNG file format (2cm × 2cm, resolutions in 300dpi)。

3.8 Source: The "Source" field is the primary key of the data system for collation.

The "Source" field consists of two data elements with one optional

element. They will be concatenated with hyphen character '-'.

- (mandatory) The 1st letter indicates the book reference number. The possible values are:

  (A) stands for《甲骨文合集》

  (B) stands for《甲骨文合集補編》

  (C) stands for《花園莊東地甲骨》

  (D) stands for《山東濟南大辛莊甲骨》

  (E) stands for《周原甲骨》

  (F) stands for《小屯南地甲骨》

  (G) stands for《英國所藏甲骨集》

  (H) stands for《 懷特氏等所藏甲骨文集》

  (I) stands for《天理大學附屬天理參考館藏甲骨文字》

  (J) stands for《德瑞荷比所藏一些甲骨錄》

  (K) stands for《瑞典斯德哥爾摩遠東古物博物館藏甲骨文字》

- (mandatory) Oracle Bone number (甲骨拓片的編號) which consists of 5 digits assigned uniquely to each Oracle Bone.

- (optional) Identifier to determine the side of the Oracle Bone which consists of 1 digit. The possible values are '0' for front side, '1' for back side. If an inscription is carved only on one side, this element will be omitted.

Three examples of the "Source" field are listed below.

- A-00001   (does not have front and back side)

- A-00001-0 (front side)

- A-00001-1 (back side)

4. The principles of Old Hanzi selection

4.1 Distinction principles

Two or more instances of Old Hanzi are considered different from each other when either one of the following conditions apply:

4.1.1 One or more types of components are different.

(莫) , (示)

4.1.2 The number of components or lines is different.

(介) , (星)

4.1.3 The direction (e.g. mirror image) of a component is different.

(人) , (師) , (至)

4.1.4 The position of one or more components is different.

(好)

4.1.5 Whether the same set of components are connected each other or not.

(伐)

4.1.6 One or more line types (straight line, curve, circle, rectangle, closed line or curve filled inside) are different.

(王)

4.2 Unification principles

Two or more instances of Old Hanzi with the following differences will be unified unless there's any difference in the meaning:

4.2.1 The length of corresponding line is different.

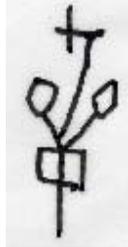**4.2.2 The thickness of corresponding line is different.**

4.2.3 The size of each component of the same set is different.

5. The principles of radical classification

5.1 The glyph should be classified into Shuowen radicals according to the Original Oracle Bone Inscriptions, e.g. "斲" should be classified under radical 斤.

歸部問題：以甲骨文字形爲主，如"斲"入斤部。

Example 1(範例一)

Rep. Script/Glyph	Original Shape/Glyph	S.W. Radical
		斤

5.2 The shapes of the Original Oracle Bone inscriptions are different, but they share the same meaning and usage.   Eventually, they have evolved into a pair of variants , with two different radicals in Shuowen.   According to the radical classification in Shuowen, the characters are put under different radicals. e.g. 兀 and 元.

甲骨文異形同用，後世分爲兩字，說文分見兩部，則依《說文解字》收入不同部首。如"元"、"兀"。

Example 2 (範例二)
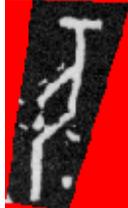
Rep. Script/Glyph	Original Shape/Glyph	S.W. Radical
		一
		儿

5.3 The shape, meaning and usage of the Original Oracle Bone inscriptions are identical but new components have been added over time.   If a radical can be found in Shuowen, the characters

will be put under the corresponding radical according to the Original Oracle Bone inscriptions. e.g. 录 put under the 录 radical and similarly 畐 in 畐 radical.

甲骨文同形同用，後世增添偏旁，《說文解字》另有部首者，則依甲骨文原形歸入相應部首。如"畐"入《說文解字》畐部，"录"入《說文解字》录部。

Example 3 (範例三)

Rep. Script/Glyph	Original Shape/Glyph	S.W. Radical
		畐
		录

5.4 The shapes of the Original Oracle Bone inscriptions are the same but they have many meanings and usages. Eventually, they have evolved into different characters. The glyph of these characters will be determined according to the shapes of the Original Oracle Bone inscriptions, and put under the corresponding radical in Shuowen. e.g. 史吏事.

甲骨文同形多用，後世分爲多字，依甲骨文字形分別隸定，歸入《說文解字》相應同一個部首。如："史"、"吏"、"事"。

Example 4(範例四)

Rep. Script/Glyph	Original Shape/Glyph	S.W. Radical
		一
		一

6. The principles of sorting the order of the glyphs of the same Oracle Bone Inscription

同字之字形排序原則

6.1 For Oracle Bone Inscriptions that are similar with Small Seal, Li style and Kai style font, they will be placed first .

甲骨文字形與小篆、隸書和楷書相合者(接近,相似),放在首位。

6.2 If one or more types of components or radical are different, then those with smaller difference will be placed first, and those with greater difference behind.

異構字依字形差異大小排序,差異較小者置於前,差異較大者置於後。

6.3 Glyph variants will be placed after the main glyph.

異寫字置於主形之後。

7. References of the Oracle Bone Inscription

The collection of the Oracle Bone Inscription will be developed to cover the following 11 books:

1. 《甲骨文合集》
2. 《甲骨文合集补编》
3. 《花园庄东地甲骨》
4. 《山东济南大辛庄甲骨》
5. 《周原甲骨》
6. 《小屯南地甲骨》
7. 《英國所藏甲骨集》
8. 《懷特氏等所藏甲骨文集》
9. 《天理大學附屬天理參考館藏甲骨文字》
10. 《德瑞荷比所藏一些甲骨錄》
11. 《 瑞典斯德哥爾摩遠東古物博物館藏甲骨文字》

8. Data Format For Old Hanzi Data Exchange

For the data exchange and review work, members are going to use the data format specified as follows:

■ **Images format:**

8.1 use PNG storage format.

8.2 The original glyph should be scanned at 300 dpi (dots per inch).

8.3 The transcribed glyph images are named [ID]+[_R] (for example, if the ID is T00001, the transcribed glyph images should be named T00001_R).

8.4 The original glyphs are named [ID]+[_O] (for example, if the ID is T00001, the original glyph should be named T00001_O).

8.5 Glyph determination images are named [ID]+[_D] (for example, if the ID is T00001, the glyph determination image should be named T00001_D).

8.6 The images of unifiable shapes are named [ID]+[Unifiable shapes ID] (for example, if the Old Hanzi ID is T00001 and the unifiable shape ID is 000, the image of unifiable shape should be named T00001_000).

■ **XML Schema:**

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
elementFormDefault="qualified">
  <xs:element name="OldHanZi">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="Character" minOccurs="0"
maxOccurs="unbounded"/>
      </xs:sequence>
      <xs:attribute name="version" type="xs:string" use="required"
fixed="1.0"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="Character">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="Source"/>
```

```xml
            <xs:element ref="Period"/>
            <xs:element ref="Area"/>
            <xs:element ref="Material"/>
            <xs:element ref="Radical"/>
            <xs:element ref="ModernChar" minOccurs="0"/>
            <xs:element ref="Unified" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="Note" minOccurs="0"/>
        </xs:sequence>
        <xs:attribute name="id" use="required">
            <xs:simpleType>
                <xs:restriction base="xs:string">
                    <xs:pattern value="(G|T|K|KP|J|V|S|H|M)[0-9]+"/>
                </xs:restriction>
            </xs:simpleType>
        </xs:attribute>
    </xs:complexType>
</xs:element>
<xs:element name="Source" type="xs:string"/>
<xs:element name="Period" type="xs:string"/>
<xs:element name="Area" type="xs:string"/>
<xs:element name="Material" type="xs:string"/>
<xs:element name="Radical">
    <xs:simpleType>
        <xs:restriction base="xs:unsignedShort">
            <xs:minInclusive value="1"/>
            <xs:maxInclusive value="540"/>
        </xs:restriction>
    </xs:simpleType>
</xs:element>
<xs:element name="ModernChar" type="xs:string"/>
<xs:element name="Unified">
    <xs:complexType>
        <xs:attribute name="id" type="xs:string" use="required"/>
    </xs:complexType>
</xs:element>
<xs:element name="Note" type="xs:string"/>
</xs:schema>
```

■ XML example:

```xml
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<OldHanZi version="1.0">
  <Character id="T00001">
    <Source>甲骨文合集</Source>
    <Period>商</Period>
    <Area>河南安陽</Area>
    <Material>甲骨</Material>
    <Radical>001</Radical>
    <ModernChar>一</ModernChar>
    <Unified id="0000"/>
    <Unified id="0001"/>
    <Note/>
  </Character>
  <Character id="T00002">
    <Source>甲骨文合集</Source>
    <Period>商</Period>
    <Area>河南安陽</Area>
    <Material>甲骨</Material>
    <Radical> 005</Radical>
    <ModernChar>王</ModernChar>
    <Unified id="0000"/>
    <Note/>
  </Character>
</OldHanZi>
```