

「2018 研究資料管理工作坊 2018/12/13

# 小學堂文字學資料庫的 資料管理與開放

莊德明 中央研究院數位文化中心研究助技師  
中央研究院歷史語言研究所研究助技師  
Email : [derming@gate.sinica.edu.tw](mailto:derming@gate.sinica.edu.tw)

# 研發概要

- \* 小學堂是形、音、義綜合的文字學資料庫。
- \* 研發單位：臺灣大學中國文學系、中央研究院歷史語言研究所、資訊科學研究所、數位文化中心。
- \* 研發時間：1993年迄今
- \* 開放日期：2013年4月26日
- \* 小學堂網址：<http://xiaoxue.iis.sinica.edu.tw>

# 收錄現況－字形

- \* 小學堂收錄的字形，包括甲骨文、金文、戰國文字、小篆及楷書，字數超過22萬個。

字形	字數	備註
甲 骨 文	24701	
金 文	23836	
戰 國 文 字	70557	其中楚系簡帛文字37614個、秦系簡牘文字3904個、傳抄古文字29039個。
小 篆	11101	含古文、籀文、篆文、或體、俗字、奇字
楷 書	91818	
	222013	

# 「員」的字形演變及解形

員 [員]

節錄自《漢語大字典》

員



乙四四三



佚一一一



員壺 (一)



或鼎



石鼓



說文籀文



說文·員部



睡虎地簡五  
三·二九



老子甲後三  
〇一



史晨碑



袁博殘碑

《說文》：“員，物數也。从貝，口聲。鼎，籀文从鼎。”林義光《文源》：“口，鼎口也。鼎口，圓象。”按：“員”是方圓的“圓”的本字。

# 員 (金文)



員鼎



員盃



員父尊  
員 从鼎



員盃



員盃  
鼎 成方



員盃



員尊

# 「員」的相關字形

	員 鼎		
	嬪 媼 勳 勳		
	隕 隕		勳 勳 勳

# 收錄現況 — 字音

- \* 小學堂收錄的字音，包括上古音、中古音、近代音、現代音，筆數超過134萬筆。

時代	筆數	備註
上古	56169	
中古	257549	含廣韻、集韻、禮部韻略等韻書
近代	28593	含中原音韻、中州音韻、洪武正韻等韻書
現代	999575	其中官話44178筆、晉語10707筆、吳語102304筆、徽語4708筆、贛語71743筆、湘語15032筆、閩語135604筆、粵語325609筆、平話94702筆、客語156544筆、其他土話38444筆。
	1341886	

# 天 (上古音)

時代	聲韻系統名稱	韻部	聲母	韻母
先秦	高本漢系統		t <sup>h</sup>	ien
先秦	王力系統	真	t <sup>h</sup>	ien
先秦	董同龢系統	真	t <sup>h</sup>	ien
先秦	周法高系統	真	t <sup>h</sup>	en
先秦	李方桂系統	真	th	in
兩漢	西漢	真		
兩漢	東漢	真		



# 數 (中古音、廣韻)

字號	字形	聲調	韻目	字母	上字	下字	又切	又音
10975	數	上	麌	生	所	矩	桑谷/所句/所角	
15851	數	去	遇	生	色	句	色矩/色角	速
20187	數	入	屋	心	桑	谷	所矩/所角/色句	
20909	數	入	覺	生	所	角	所矩/色句/桑谷	

# 收錄現況—索引

\* 小學堂收錄的字書索引，筆數超過36萬筆。

類別	字書
甲骨文	甲骨文編、甲骨文字詁林、甲骨文字集釋、殷墟甲骨刻辭類纂、新甲骨文編、甲骨文字編
金文	金文編、金文詁林、金文詁林補、殷周金文集成引得、新金文編
戰國文字	楚系簡帛文字編、睡虎地秦簡文字編、秦簡牘文字編、戰國古文字典：戰國古文聲系、傳抄古文字編
小篆	說文解字詁林正補合編、說文新證
隸書	秦漢魏晉篆隸字形表
楷書	漢語大字典、康熙字典、中文大辭典
韻書	廣韻、集韻

# 「直」字索引(金文、甲骨文)

- 金文編（中華書局），頁.行：833.6
- 金文詁林補（中央研究院），冊.卷.頁：6.12.3692
- 殷周金文集成引得（中華書局），字號.頁：4330.1275
- 新金文編（作家出版社），冊頁：中1741
- 甲骨文編（中華書局），頁.行：497.4
- 甲骨文字詁林（吉林大學），冊.頁：1.554
- 甲骨文字集釋（中央研究院），卷.頁：補遺.4438
- 殷墟甲骨刻辭類纂（吉林大學），冊頁：上206
- 新甲骨文編（福建人民出版社），頁.行：699.4
- 甲骨文字編（中華書局），字號.冊頁：0652.上185

# 連結的線上字典

\* 小學堂已連結四部線上字典。

線上字典	負責單位	網址
異體字字典	教育部開發，現由國家教育研究院維護	<a href="http://dict.variants.moe.edu.tw/">http://dict.variants.moe.edu.tw/</a>
CNS11643中文標準交換碼全字庫	原行政院主計處電子處理資料中心開發，現由國家發展委員會維護。	<a href="http://www.cns11643.gov.tw/">http://www.cns11643.gov.tw/</a>
開放康熙字典	中華開放古籍協會	<a href="http://kangxi.adcs.org.tw/kangxizidian/">http://kangxi.adcs.org.tw/kangxizidian/</a>
國際電腦漢字與異體字知識庫	中央研究院	<a href="http://chardb.iis.sinica.edu.tw/">http://chardb.iis.sinica.edu.tw/</a>

# 直 (形音義連結)

1034

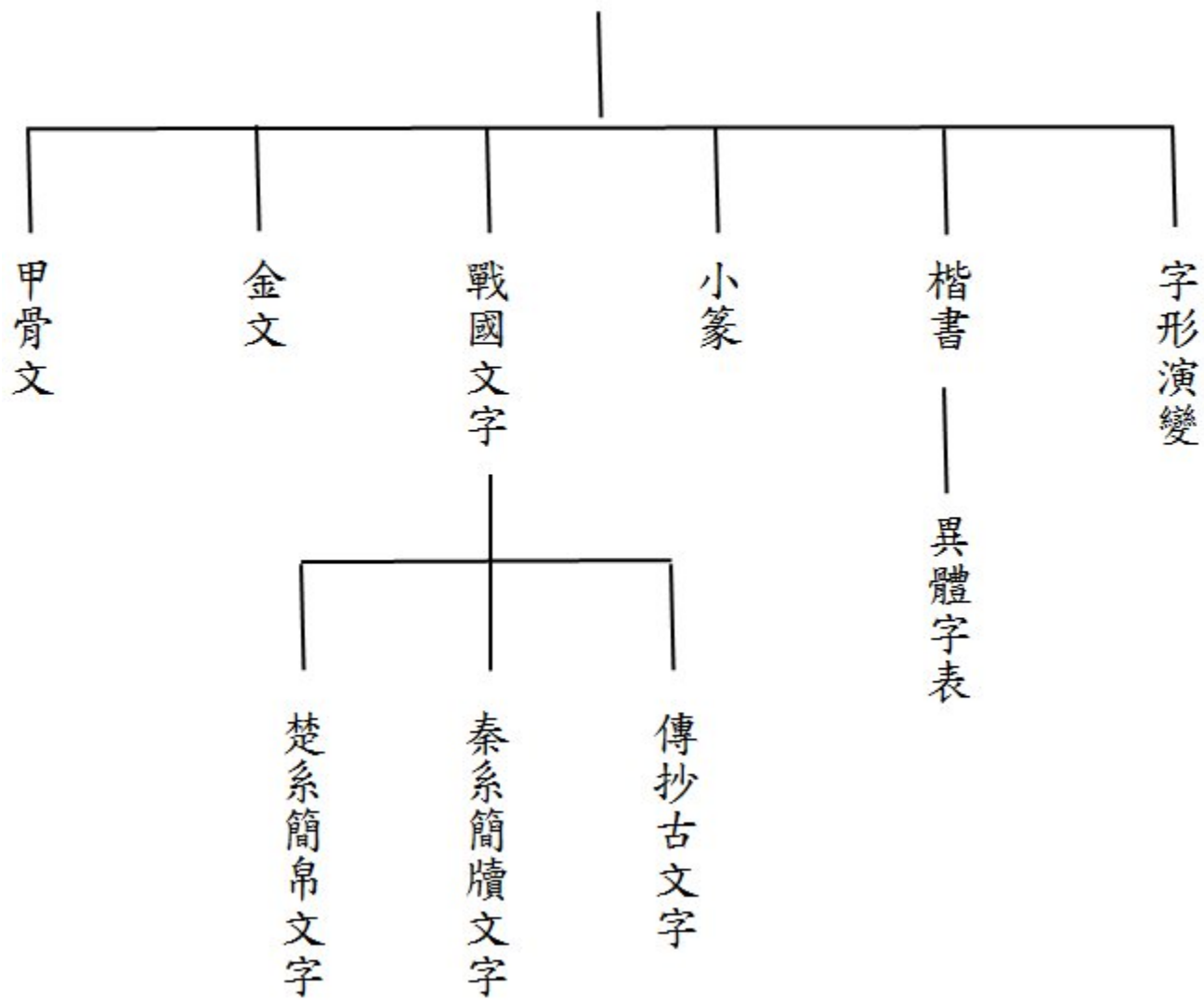
陽藏

## 直

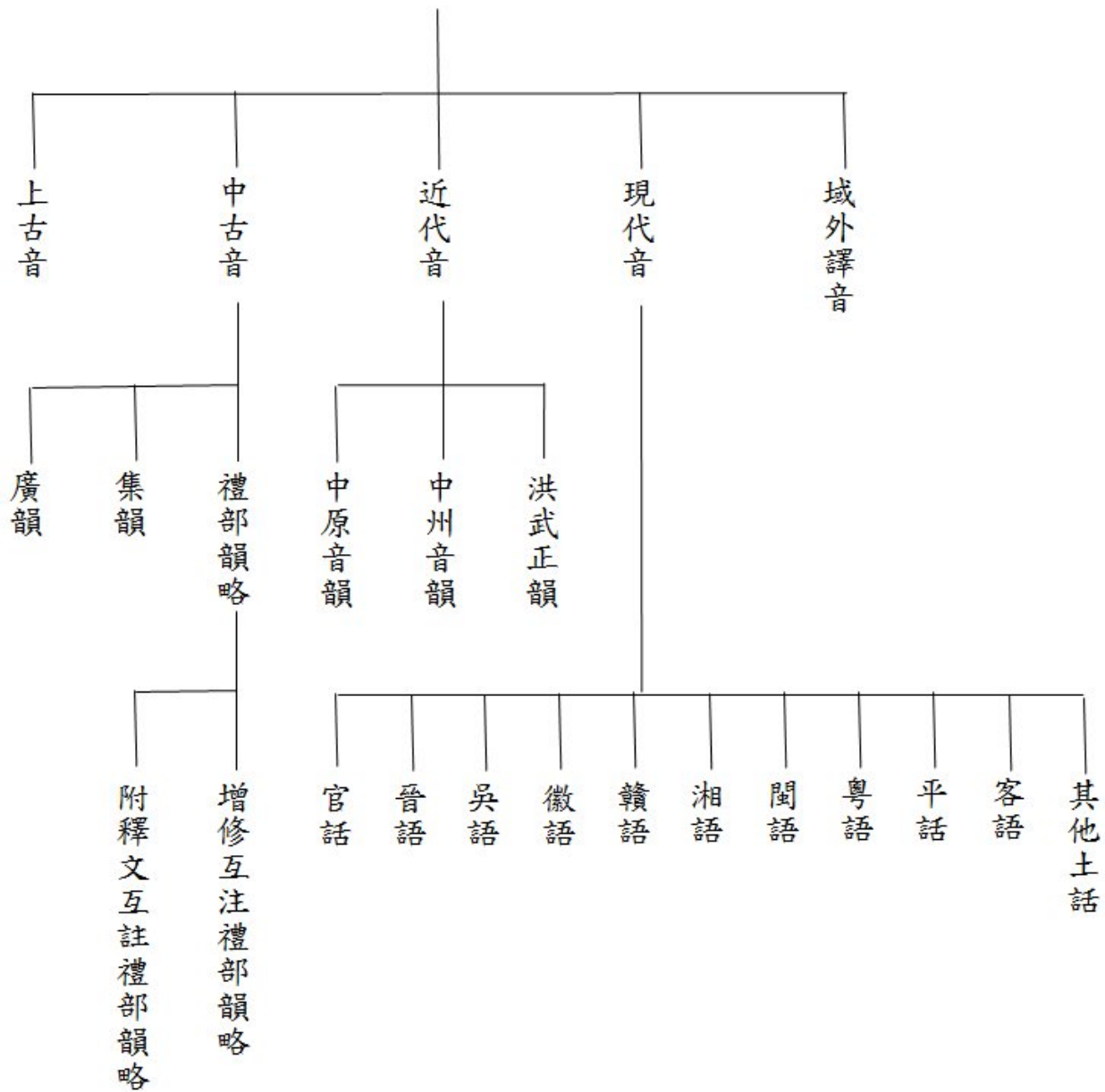
### 相關連結

- [小學堂異體字表](#)、[字形演變](#)、[小篆](#)、[金文](#)、[甲骨文](#)、[楚系簡帛文字](#)、[秦系簡牘文字](#)、[傳抄古文字](#)
- [小學堂上古音](#)、[中古音](#)、[官話](#)、[晉語](#)、[吳語](#)、[徽語](#)、[贛語](#)、[湘語](#)、[閩語](#)、[粵語](#)、[平話](#)、[客語](#)、[其](#)
- [小學堂韻書集成](#)、[廣韻](#)、[集韻](#)、[附釋文互註禮部韻略](#)、[增修互注禮部韻略](#)、[中原音韻](#)、[中州音韻](#)
- [國際電腦漢字與異體字知識庫](#)
- [教育部異體字字典](#)、[常用國字標準字體筆順](#)
- [CNS11643 中文全字庫](#)
- [開放康熙字典](#)

# 漢字古今字資料庫



漢字古今音資料庫



# 網站流量分析

2013年4月26日 - 2018年12月3日

所有使用者  
100.00% 個使用者

+ 新增區隔

總覽

瀏覽量 ▾ 對比 選取指標

每小時 天 週 月

● 瀏覽量



使用者  
875,709

新使用者  
832,437

工作階段  
1,643,677

每個使用者的工作階段數量  
1.88

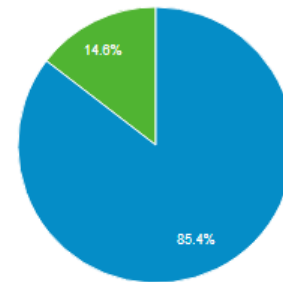
瀏覽量  
13,908,302

單次工作階段頁數  
8.46

平均工作階段時間長度  
00:08:15

跳出率  
21.04%

■ New Visitor ■ Returning Visitor





小學堂文字學資料庫運作兩年來，解救了無數文字學研究者的惡夢……過去我們為了找一個字，要在一套一套大部頭的工具書間查找，自行跨工具書整合所有查詢結果，光是『查詢』這個工作就要耗掉半天，現在則縮短到了一秒之間。」

陳穎青（2015）。甲骨文研究的數位福音  
科學人2015年第163期09月號

甲骨、青銅器、竹簡木牘是千百年前的物品，經過這麼長的時間，會壞掉、鏽蝕，就算是從拓本或文字編剪字，要把這麼多字一個一個在網站上呈現的黑白分明，一定是很花時間很辛苦的事情！從「漢字構形資料庫」軟體開始，就覺得使用這樣的軟體好奢侈啊，字形這麼漂亮、解析度這麼高，按個ctrl+w就可以完美的鑲進word裡！小學堂網站也是會讓使用者覺得好奢侈的網站！

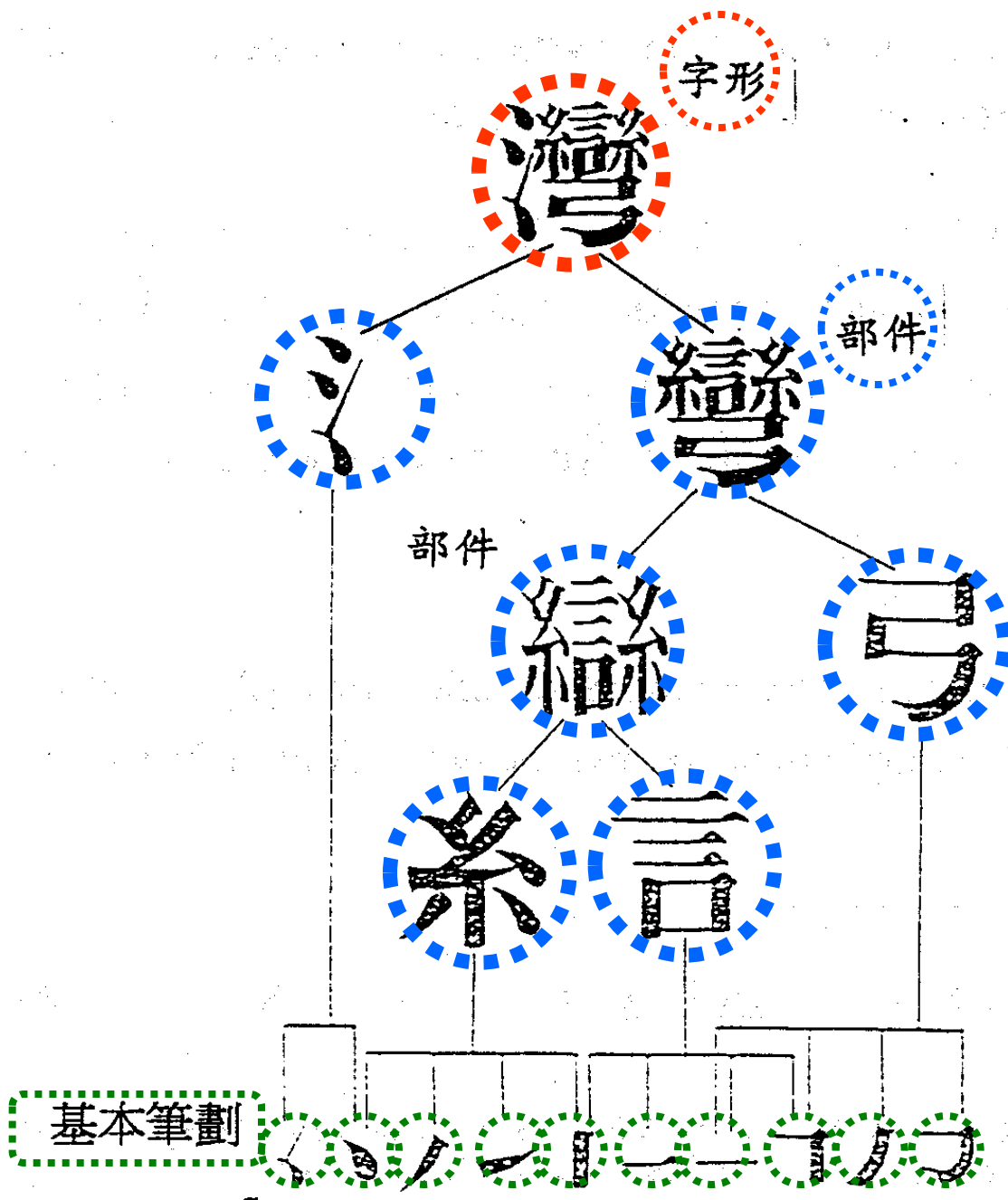
小學堂使用者 2016/7

正確的漢字演變知識不必來到中文系才學會、才懂得，只要在中學任教的國文老師懂得如何善用小學堂，我們的學生可以在中學就建立很好的漢字演變觀念。

國立暨南國際大學中文系陳美蘭老師 2016/7/27

小學堂資料庫其覆蓋的內容是多麼完整、其製作的過程是多麼嚴謹！這是一個足以傳世之作！這個資料庫之所以取得如今的成就，實在是由於許許多多學者、工程師全心全意投入，合作無間，為著一份使命感、一個理想，不較個人的利害得失，憑著堅忍的毅力，用自己的血汗，一步一步的完成的；因為，這個工作並沒有一個從頭到尾的大型計劃支持，是一路摸索、逢山開路遇水搭橋地走過來的。

中研院資訊所兼任研究員謝清俊 2017/10/10



# 小學堂資料庫建置大事記

- \* 1993年資訊所文獻處理實驗室開始重建「交大字根系系統」，此系統為交通大學於1972年開發。
- \* 1995年4月，資訊所謝清俊教授發表〈中文字形資料庫的設計與應用〉一文，此資料庫記錄了8,532個字形結構，並提供系統磁片與大家分享。自此以後，最新的資料庫成果即以磁片、光碟，或提供下載網址來分享。
- \* 1996年8月，謝清俊教授發表〈電子古籍中的缺字問題〉一文，提出利用字形結構表達式（簡稱構字式）來解決缺字問題。

# 小學堂資料庫建置大事記(續)

- \* 1997年，臺灣大學中國文學系楊秀芳老師開始建置漢字古今音資料庫。
- \* 2002年5月，資訊所文獻處理實驗室推出漢字構形資料庫1.4版，收錄《漢語大字典》字頭，合計54,711個字形（含結構）；另收錄《漢語大字典》異體字表12,208組。
- \* 2003年3月，漢字構形資料庫更新到2.1版，完整地收錄了《說文解字詁林》的小篆及重文字形（含結構）11,000個，正式跨入了古漢字的領域。

# 小學堂資料庫建置大事記(續)

- \* 2003年4月到2007年8月，在史語所文字學門的協助下，漢字構形資料庫陸續收錄金文21,413個（含結構）、楚系簡帛文字19,138個（含結構）、甲骨文2,197（含結構）。
- \* 2009年7月，楊秀芳老師邀請資訊所文獻處理實驗室共同開發「漢字古今音資料庫」。
- \* 2011年9月，漢字古今音資料庫正式開放給外界使用。
- \* 2013年4月，漢字構形資料庫隨後依循漢字古今音資料庫的架構，重新開發提供線上查詢，並整合成小學堂文字學資料庫供外界使用。



# 小學堂資料庫建置大事記(續)

- \* 2013年9月，小學堂文字學資料庫改由數位文化中心維運。
- \* 2013年9月至2016年，陸續加入字形演變，並擴充文字及聲韻資料，其中包括戰國文字32,943個。

# 遭遇過的困難

- \* 字碼 (Big5, Unicode CJK/Ext-A/Ext-B/Ext-C/Ext-D/Ext-E)
- \* 作業系統 (Windows 3.1/95/98/2000/xp/7/10)
- \* 開發環境 (Windows/Web)
- \* 開發工具 (VB 3.0/6.0/.net)
- \* 引用資料的智財權爭議 (創用 CC 姓名標示)

謝謝聆聽