

數墓：台灣墓碑數位為典藏、教學、及研究 數位化工作流程

計畫單位：國立高雄大學西洋語文學系

計畫名稱：「數墓：台灣墓碑數位為典藏、教學、及研究」

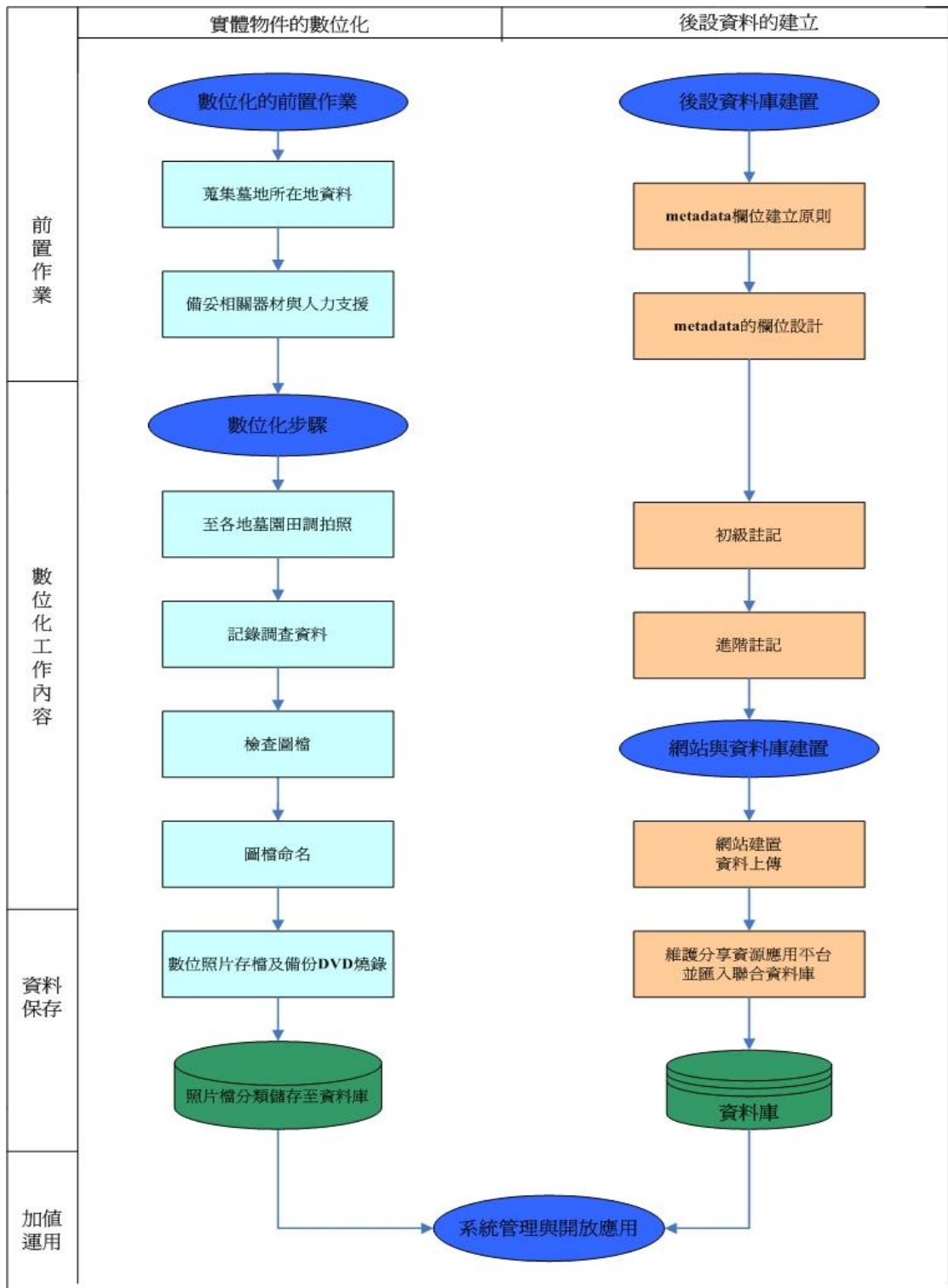
計畫簡介：

墓碑除代表個人的記憶認同，也能整體的顯現社會的認同、語言、習俗與工藝的傳承。然而基於社會與氣候因素，台灣當今的墓碑卻戲劇性地大量消失；為了保存墓碑上所顯現的豐富文化遺產，我們於 2007 年開始了 Thakbong 的計畫，其計畫主旨為將台灣豐富的墓碑建檔數位化，其資料的保存為跨族群、跨宗教、跨區域、跨階級，甚至是橫跨台灣歷史年代的。此數位建檔基本上使用地理資訊系統以保留其地理位置資訊，以及 TEI 相同的 XML 系統將其墓碑刻字以及墳墓狀況以文字註解建檔成資料庫，而此資料庫將做為研究、教學，以及典藏用途。

此研究計畫目標為透過電腦化分析、數據的觀察，來探討台灣人的文化、自我認同。我們將建立一個以三萬筆台灣墓碑資料為基礎的語料庫，其涵蓋年代跨越許多時期從 1800 年到 2005 年。此研究除了針對人類語言學的議題討論，也跨足許多其他領域，包含電腦語言學、語料庫語言學、數位資料保存以及心理語言學。

計畫網站：<http://thakbong.dyndns.tv>

數位化工作流程圖



數位化工作流程說明

一、前置作業

必須先收集墓地所在地資訊，例如地圖上找有墓地記號者(倒“T”字型)，或者用 Google earth 也看得到。此外，也有一些小的墓地，可能因為沒有名字，也可能是非法的，或是亂葬崗之類的，都是普查的調查範圍。一些非公家管理的墓地只要隨機看到就拍，有機會出公務或遊玩就拍(至今已拍了八萬多筆照片)。田野調查的主要硬體工具包括有：GPS 裝置之數位相機共 5 台(SONY Cyber-shot DSC-HX5 1 台、RICOH G700SE 2 台、Samsung HZ35W 2 台)。必要時還需備有除草工具。因田調作業受氣候影響，下雨天或晚上不利調查與拍攝就不會前往。

唯某些季節的天候狀況不穩定，田調的日期很難預先規劃。已荒廢或疏於管理的墓必須由計畫人員除草及清理後方能拍攝。地層下陷、鄉鎮及都市更新計畫都可能造成墓園的損害及搬遷。部分地區亦可能潛藏蛇、蜂、有毒植物等危險生物。工作進度則依氣候條件、日照充足等情況有所不同，每次田調從數小時至一整天不等。



圖 1：地圖上倒 T 字型記號為墓地所在



圖 2：相關工具（手套、粉筆、除草用具等）



圖 3：拍攝用的所有相關器材，包括內建 GPS 定位系統的相機

二、數位化步驟

（一）至各地墓園田調拍照（Field work, taking photos）

每座墳墓含整體結構、墓碑、供品、土地公等在內拍攝 2 至 7 張照片(依墓的大小而定)，其標準作業流程由計畫主持人訂定。不同墓園及不同墳墓之邊界應以一張額外拍攝的交通工具(汽、機車)照片作為區隔。

其主要的拍照程序有：

1. 大致環視一下週遭環境，觀察附近的狀況，並決定如何分工進行調查拍攝。



圖 4：觀察一下環境（左：黃駿，右：奧利華教授）

2.以汽車作為邊界區隔的物件，拍攝照片以利比對。



圖 5：拍攝汽機車邊界以做為區隔

3. 拍攝的順序以大環境一張，接著是正面墓地，再拍攝墓碑左方、右方和土地公（后土）、石頭、祭祀物等物件，以及拍下一個墓碑作為相對位置比對，最後再以粉筆做記號表示已調查、拍攝過。（圖 6～圖 10）有傾倒的碑，需搬動翻正、清除塵土再拍，拍過後一樣需復原並做記號。（圖 11）
- 4.



圖 6：拍攝墓碑的左側



圖 7：拍攝墓碑的右側



圖 8：拍攝土地公（后土）



圖 9：拍攝墓碑上的紙錢或石頭等物件



圖 10：每完成一個墓碑拍攝以粉筆做記號，避免重複拍攝



圖 11：傾倒的碑，需搬動翻正、清除塵土再拍

（二）記錄調查資料

除了拍攝墓碑（地）數張外，主要還得採集記錄相關資訊，包括：type, title, subject, source, rights, relation, publisher, language, identifier, format, description, coverage, contributor、宗教外，以及當時環境有無的樣貌，例如山林、日本墓碑、土地公、天主教、金爐等數十種的資訊調查記錄。有時還會依據基地的環境給予等級的評比，做為研究參考數據。此類的訊息都會記載在制式的表格內。

因田調地處不便居多，加上拍攝時間有限。有時記錄表格若無法當場填寫，多半會回到工作室時盡早填寫。加上目前所使用的相機都有定位記錄系統，可有效記載照片的方位地理資訊。早期則是拍照和 GPS 系統分別攜帶拍攝，較為不

便。



圖 12：早期拍攝的方式都是照片和 GPS 分別拍攝記錄

三、數位照片存檔及備份 DVD 燒錄 (Saving the photos into hard disks and DVD)

田調後的拍攝原始圖檔都會由主持人、博士後助理及研究生助理共 3 名人員先儲存於備份硬碟。每批田調的照片至少儲存於兩台硬碟及一張 DVD 中共三處。依每批田調所得照片數量多寡，下載及燒錄所費時間亦不一。



圖 13：儲存資料硬碟與光碟

四、將照片檔分類儲存至計畫資料庫中，並上傳至網路介面以供助理註記 (Sorting the photos into database)

這些原始 jpeg 圖檔約為 4MB，經確認經緯度及方位資料已寫入，且墓碑文字及圖像符號清晰可見，再轉為約 2MB 大小(jpeg)。確定墓與墓的邊界後，將同一座墓的照片納入同一個影像群組，再上傳至透過 XHTML 所建立的網路介面上，供註記員上網註記細節。

因下載速度以及程式所能處理的資訊之複雜程度皆會影響工作進度、效率。

其中資料處理程式均由主持人撰寫，並不定期視需要隨時修正。這些照片大部分之後會上傳至聯合目錄，但畫面拍的不清楚者就不傳，照片無版權問題的也才能送至公開網路。

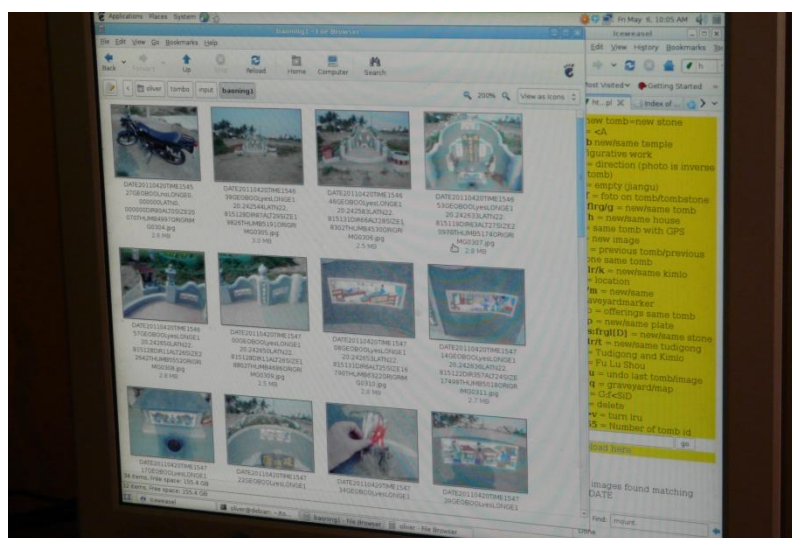


圖 14：同一座墓的照片歸為同一群組，再上傳至網路供工作人員做細節註記。

五、 後設資料建檔流程

(一) 後設資料建立 (Metadata)

由主持人先依計畫之需求，以英文建立 metadata，再由助理翻譯成中文，或由臨時工翻譯成其他語言(阿拉伯文，日文)。助理另需與 TELDAP 小組協調，並視需要重新整理 metadata。

(二) 初級註記 (Basic Transcription and Annotation)

主持人透過 XHTML 建立註記介面，作為基本的標準。註記人員由網路進入該介面，將墓碑上的文字打字記錄下來，並經由下拉式選單勾選基礎的 metadata 資料。註記人員若有問題，例如說墓碑上的字跡難以辨視或該字為難以打字的異體字時，可點選“problem”，將該筆資料回報給進階註記人員處理。Metadata 目前共有以下 12 分項：type, title, subject, source, rights, relation, publisher, language, identifier, format, description, coverage, contributor. 其網路資料庫上的後設資料主要欄位內容如下範例：

表 1：後設資料欄位 (範例一)

type	靜態圖像
title	台灣，嘉義縣，太保市，玉山莊，近國1號嘉義交流道，無名氏之墓碑
subject	* 姓名:無名. * 姓名(拼音):WU2MING2. * 性別:不明確的. * 種類:全墓.
source	* 數位照相機: RICOH Caplio 500SE. * 田野調查: 2008-06-29. * 全球定位系統: RICOH Caplio 500SE.
rights	* 統一資源定位符:從2010-12-31以來姓名標示-非商業性-禁止改作 2.5通用版 (CCBY-NC-ND 2.5). * 公有領域:2058-12-31 以後.
relation	* 墓園: http://thakbong.dyndns.tv/graveyard/123 . * 墓碑: http://thakbong.dyndns.tv/stone/11369 . * 墓: http://thakbong.dyndns.tv/tomb/11143 .
publisher	國立高雄大學
language	* 文字:中文. * 口語:可能是河洛語. * 書寫方向:由上到下,由右到左.
identifier	* 檔案: thakbong/graveyards/123/123-2008-06-29/DATE20080629TIME165902GEOBOOLyesLONG120.373103LAT23.498133ALT25.jpg. * 照片的的統一資源定位符: http://thakbong.dyndns.tv/image/26377 . * 墓碑的統一資源定位符: http://thakbong.dyndns.tv/stone/11369 . * 數位典藏與數位學習國家型科技計畫的統一資源定位符: 2809.
format	* 資料壓縮:無. * 網際網路媒體型式: image/jpeg * 大小:1084483 位元組.
description	這是一個在台灣，嘉義縣，太保市，玉山莊，近國1號嘉義交流道發現的墳墓。這個墓碑上的文字都是直接鑿刻上去的。這個墓碑的形狀是柱形。這個墓碑上所記載的訊息包括「種族別」。
coverage	* 定位:東23.498133 北120.373103. * 時代:中華民國,二戰後(1945-2008).
contributor	* 資料註記及建檔:奧利華. * 資訊管理:奧利華. * 田野調查:奧利華. * 品質管理:黃駿.

資料庫後台的初階註記，包括了基本資料 male、child、female、family name、religion 等數十類欄位。每一組照片檔案的描述還有簡易的縮寫代號，例如：Q 公墓、e 外面、D 正面、i 裏面、O 祭品供品、T 土地公、r 右邊、k 金爐、f 左邊等代號。加上根據多年的經驗已經建立了一套預設資料選單，每一欄位只要選擇正確的內容填上即可，若遇到選單沒有的資料才會再增加補上。初階註記完成後，才會上傳至下一階段。（圖 15～圖 16）



圖 15：資料庫後台的初階註記欄位-1



圖 16：資料庫後台的初階註記欄位-2

(三) 進階註記 (Advanced Transcription and Annotation)

檢視通過初級註記的資料，處理初級註記員回報的問題，確定沒有問題後將該墓碑影像及註記資料送入計畫網站。主持人及博士後助理共 2 名人員能鑑別初級註記後的資料是否符合標準，亦能處理初級註記人員所回報的問題。在該筆資料確認無虞後，即已達到可匯入聯合資料庫之標準。



圖 17：進階註記工作

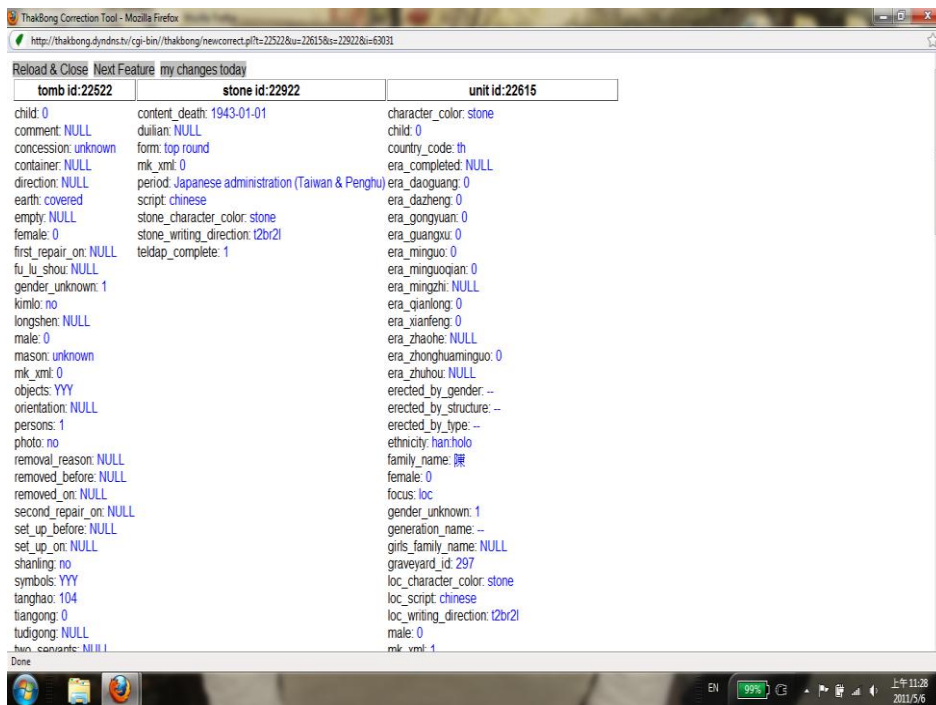


圖 18：註記後的所有欄位一覽

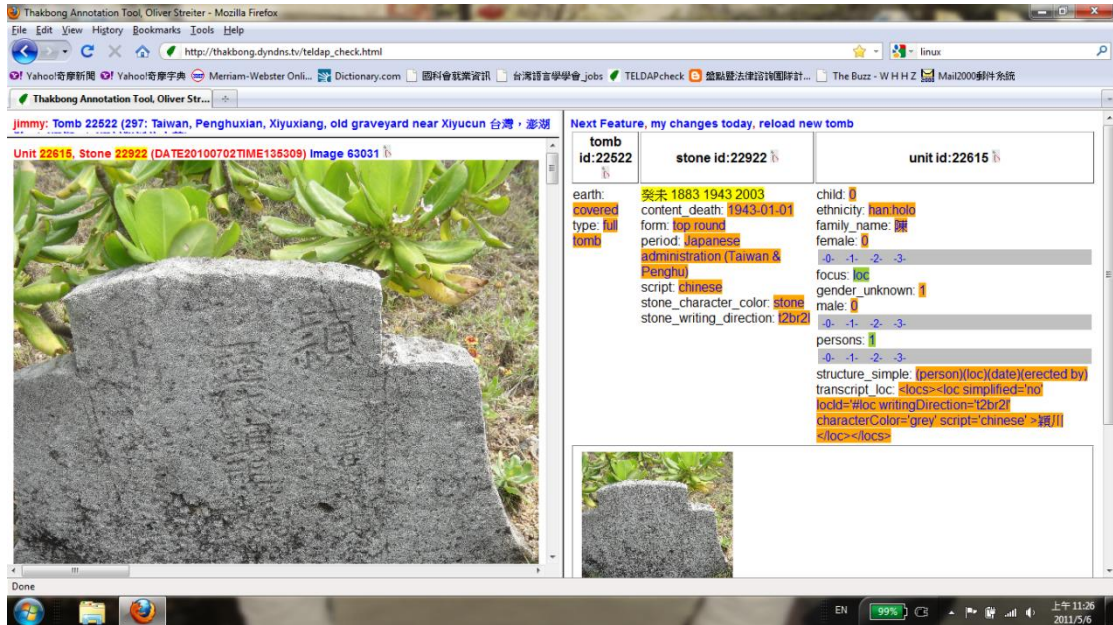


圖 19：持續進階註記工作

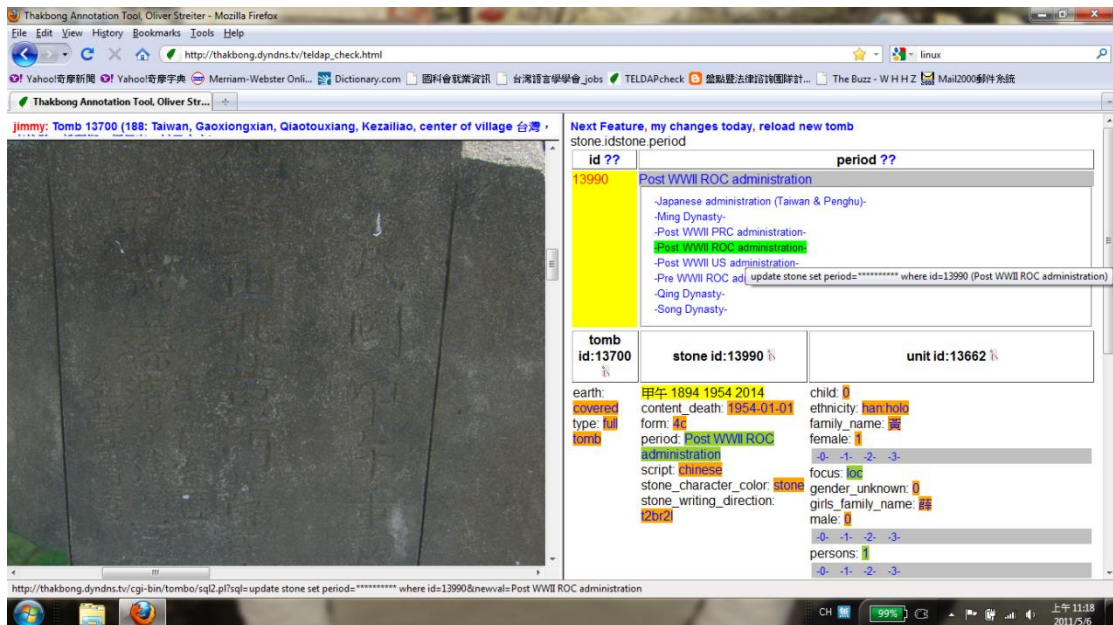


圖 20：進階註記工作，年代確認的標記、墓碑上的文字描述等範例

六、數位化成果資料庫建置

如上段所述，在訂定後設資料欄位時即考量配合資料庫建置與使用者便利性，建立瀏覽查詢、資料庫搜尋等機制，方便不同查詢需求的閱覽者進行檢索，資料庫內容亦公開分享予教學、研究各相關社會大眾廣泛使用。查詢欄位大致有依台灣各地縣市、姓氏等方式查詢，還結合 google earth 可直接在地圖上點選資

料 (圖 21)。



圖 21：結合 google earth

七、計畫網站製作、維護，並將註記好的資料上傳 (Releasing data onto the project's website)

由主持人負責網站架設，兩名助理則負責英文漢譯及漢文版面的維護。依循網路通訊協定 (Communications Protocol) 之標準，網站目前為英文漢文雙語。(<http://thakbong.dyndns.tv/>) 網站架設後由裝置每日不間斷地自動進行。主持人另於每筆新資料註記完成後上傳，所需時間依檔案大小不同。每個月至少上傳 1,000 筆墓碑資料。

ThakBong: Digitalizing Taiwan's Tombstones for Teaching, Research and Documentation

Collaborators: Oliver STREITER, Yoann GOUDIN, Ann Mefang LIN, Jimmy HUANG Sandy Ko-ju YEN

What is ThakBong?
ThakBong is an archive of tombstones in Taiwan and some other places where Chinese settlers have moved to. The archive is realized as an XML-structure which links to images taken from gravesites in Taiwan, China, Hong Kong, Hawaii, USA etc.

What is DigiBong?
On 2010.08.01 we got sponsorship for one year for our activities in the framework of TELDAP under the grant number 99-2631-H-390-002. In this framework, the project is called DigiBong standing for the 'Digitization of Tombs'. For one year we will do now more archiving and documentation than research on tombstones. Here the DigiBong logo.

ThakBong Tombs in Taibeishi

Taiwan, Taibeishi, Nangangu, second graveyard

photos	tombs	visits	period	researcher
144	28	2	1971-01-01 - 1990-01-01	Other

most frequent family name

family name	family names	most frequent maiden name	Mainlander
CHEN2	CHEN2 (2), LI3 (1), CAH4 (1), LIN2 (1), HUANG2 (1), JIAO1 (1), ZHAN1 (1), LIAO4 (1)	WANG2	3

Tomb: 13405
Google Maps Google Earth
Date: 2008-12-27, Time: 062416

Search Tombstones through Names

Find tombstones matching a family name: AH, AH1, BA1, BA2, BAO1, BAO3, BEN1, BO2, BU3, BU4, CAH, CAO2, CEN2, CHEN2, CHENZHONG2, CHENG2, CHENG2, CHU2, CHU2, CHU2, CHU3, CU1, DAI1, DENG4, DIAO1, DING1, DONG1, DONG2, DONG3, DOU4, DU4, DUAN1, DUAN2, DUO1, FA3, FAN2, FAN4, FANG1, FANG4, FEI3, FEH, FENG1, FENG1, FENG2, FENGA, FU2, FU2, FU3, FU4, GAO1, GE3, GONG1, GONG1, GU1, GU3, GU4, GUAN1, GUAN1, GUANG, GUANG3, GUO1, HAN2, HE2, HE3, HEI, HAO3, HONG2, HONG2, HOU2, HOU4, HOU4, HUB, HU3, HUAT, HUAT, HUANG2, JI2, JI3, JIAN2, JIANG1, JIANG1, JIANG3, JIAO1, JIU1, JIU1, JIANG1, KE1, KONG3, KUANG4, LAH4, LAN2, LAN2, LAN2, LAO2, LEI2, LI2, LI3, LI3LIN2, LI4, LI4, LIAN2, LIAN4, LIANG2, LIAO4, LIN2

圖 22：網頁資訊

八、維護分享資源應用平台並匯入聯合資料庫（Merging into TELDAP inventory）

維護此分享資源平台，並整理多筆符合聯合目錄規格的檔案照片上傳匯入該資料庫。因有部分墓碑文字不清晰或後設資料待考者，則暫緩匯入，盡量以完整的資訊先與大眾運用，未來將陸續開放更多筆的資料。

- ※ **製作單位：**國立高雄大學西洋語文學系「數墓：台灣墓碑數位為典藏、教學、及研究」計畫
數位典藏與數位學習國家型科技計畫 拓展臺灣數位典藏計畫
- ※ **文字撰寫：**國立高雄大學「數墓：台灣墓碑數位為典藏、教學、及研究」計畫
主持人奧利華(Oliver Streiter)老師、計畫助理 黃駿
數位典藏與數位學習國家型科技計畫 拓展臺灣數位典藏計畫
計畫助理 王雅萍
- ※ **圖片拍攝：**數位典藏與數位學習國家型科技計畫 拓展臺灣數位典藏計畫
計畫助理 王雅萍、高朗軒
- ※ **圖文編輯：**國立高雄大學「數墓：台灣墓碑數位為典藏、教學、及研究」計畫
計畫助理 黃駿
數位典藏與數位學習國家型科技計畫 拓展臺灣數位典藏計畫
生活與文化主題小組 王雅萍

致謝：

感謝國立高雄大學西洋語文學系「數墓：台灣墓碑數位為典藏、教學、及研究」計畫主持人 Oliver Streiter 老師、以及計畫助理黃駿先生、林美芳小姐等相關計畫工作人員撥冗指導及熱心協助拍攝與編寫。