

原住民族委員會
「原住民族語言研究發展中心」
(案號：106154)

附件九：數位族語資料庫
第一期建置內容及規劃修正報告

計畫執行期間：107年1月1日至107年12月31日止
(最後修改：108年1月30日)

Akiw 徐中文 研究員

Kacaw 黃堂維 研究助理

原住民族語言研究發展中心
國立台灣大學語言學研究所

本文件屬原住民族語言研究發展中心版權所有

目錄

壹、前言	1
貳、數位族語資料庫的概念	3
一、「數位族語資料庫」的想像	3
二、現況分析：參考國內外語言資料庫與語言典藏相關網站	4
參、數位族語資料庫計畫執行內容	17
一、數位族語資料庫建置目的	17
二、計畫執行方式	17
(一) 族語資料及文獻調查	18
(二) 田野訪談	19
(三) 研商會議	20
三、數位族語資料庫系統建置內容	24
(一) 資料庫基本頁面	24
(二) 資料庫系統功能(後臺)	43
(三) 已建置之資料	51
四、計畫執行成效	52
肆、數位族語資料庫修正規劃	53
一、資料庫建置規劃修正	53
(一) 增加各族語言介面切換功能	53
(二) 典藏資料功能調整	53
(三) 族語新聞功能	53
(四) 次網頁檢視	53
(五) 會員功能	54
(六) 語料建置	54
(七) 維運備份	54
(八) 資料庫內容架構	54
(九) 智慧財產權與授權	55
二、階段性建置任務及目標	56
伍、結論與建議	58
一、2019年工作規劃	58
二、其他建議	60
陸、參考文獻及網站	61
附錄一、原住民族語言研究發展中心數位族語資料庫首頁介面翻譯清單	66
附錄二、諮詢訪談及重要會議相關	67
附錄三、計畫相關會議簽到統計及會議記錄	68

「數位族語資料庫第一期建置內容及規劃修正報告」

Akiw 徐中文 研究員

壹、前言

聯合國教科文組織（UNESCO）2009 年公布了「世界瀕危語言地圖」，其中臺灣有 5 種語言被列入極度瀕危，10 種語言被列入滅絕等級。依原民會公告之「107 年度原住民族瀕危語言復振補助計畫」，列入瀕危語言的語別有卑南語、撒奇萊雅語、噶瑪蘭語、邵語、拉阿魯哇語、卡那卡那富語、賽夏語、茂林魯凱語、萬山魯凱語、多納魯凱語。

因此除了振興語言之外，亦需要積極的保存語言，正如 2016 年 6 月公布實施的原住民族語言發展法，也將「建置原住民族語言資料庫」視為一項重要的工作。Crystal（2010:349）提到保存語言的方式，即是從語言記錄開始，以錄音（聲音與影像）、分析、文字記錄來保存語言，也能夠用於教育。而當語料被記錄後，我們需要有一個整合性的平臺來收錄這些語料，整理、分類後，讓更多人能夠再應用，達到語料保存的目的。

我們查詢國內外族語保存相關的資料庫發現，國外早已有像是 AILLA、PARADISEC 這種網站，平臺提供收集多媒體語言資料及檔案，一般人都可以將所持有的媒體語料提供給該站，就會以數位化的方式被建置，甚至有分級訪問權限的制度，因此吸引許多研究人員或是族人願意提供資料。而這些資料也歡迎其他研究者協助分析、轉譯，讓資料庫收錄的越來越豐富。反觀國內，目前尚未有這種形式的資料庫，因此這也是我們努力的方向。

從歷史的角度來檢視，原住民族的語言開始有文字的記錄約莫是從荷據時期的「新港文書」開始，到日據時期之後有多位日本語言學者如小川尚義、淺井惠倫、土田滋、森口恒一、月田尚美、野島本泰為臺灣南島語言研究貢獻良多（李壬癸，2009）。而民國 57 年基督教聖經公會也以羅馬拼音將聖經翻譯成族語版，民國 1992 年則由教育部出版，委託由中央研究院李壬癸教授編訂臺灣南島語言的語音符號系統。臺灣原住民族語言雖因

口傳方式保存不易，所幸尚有外來的傳教士、學者的記錄，使許多珍貴的語料尚得留存至今，但由於早年影音記錄及保存的技術尚不發達，可能造成部分資料因此損壞或遺失，十分可惜。爰此，如何應用現代技術，將已留存的族語文獻資料，特別是將早期的族語資料彙整，並以數位化的方式保存，是首要之務。將影音的部分加以轉譯成文字，以利後續研究或教學使用，便是我們需要努力的課題。

本年度我們主要的工作是依 2017 年「原住民族語言研究發展中心數位族語資料庫建置評析報告」中的規劃，開始建置第一期數位族語資料庫系統，包含基本的網頁頁面、系統後臺、資料典藏、族語新聞、檢索功能、原住民族語言研究書目等功能。另一方面，我們也開始著手將收集的資料數位化，以及請族人協助語新聞轉譯編輯工作，並將這些資料建置於系統中，目前我們有 2 篇敘事訪談、5 則族語新聞、10 筆長篇語料典藏。目前經過測試，及經「族語資料庫修正研商會議」後，我們會再做系統的調整，期望數位族語資料庫能做為「財團法人原住民族語言研究發展基金會」典藏原住民族語料及建置原住民族語言資料庫工作的基礎。

貳、數位族語資料庫的概念

一、「數位族語資料庫」的想像

在進入正式的計畫討論之前，我們先談談族語「資料庫」和「語料庫」的差別，這也是我們在規劃的過程中一直在思考的問題。簡單的來說，資料庫 (DataBase) 其實就是儲存資料的地方，而較正式的定義則是：資料庫是由一群相關資料的集合體 (李春雄, 2014)。而藉由資料庫管理系統 (DataBase Management System, DBMS) 提供的語法功能來管理資料庫裡的資料，而透過資料庫系統的建置，使用者就能檢索、搜尋所需資料。Crystal (1991:117) 語料庫 (corpus) 的觀點來看，他認為語料庫是對語言資料收集，資料來源可以是書面語或是口語的轉譯文本，它們可以被當做語言描述的基礎或驗證關於語言假設的方法。而陳浩然、潘依婷 (2017:2) 則提到語料庫是指透過大量且根據特定原則進行文本搜集，以電子格式儲存，也就是使用者可以將大量的語言使用情形，進行詞彙、句子的檢索和統計。而語料庫中所收集的語料通常具有真實性、代表性、規模性、結構性、可機讀性 (何婷婷, 2003:4)，即可知語料庫和資料庫的最大差異在於語料庫所使用的語料文本是曾經處理，甚至分析及標記，例如收到音檔，要先轉成文字再進而分析。資料庫則是收到資料，進行數位化，輸入描述資訊檔編碼即儲存，未有語料標記及分析的工作。而現在也有很多技術，如 OCR 光學字元識別 (OCR, Optical Character Recognition)，可以幫助書面資料數位化後的字型辨識，再由人工確認，不必從頭開始打字，增加很多效率，但辨識度還是要視資料狀態而定 (程婉如, 2008)。

因此，我們認為「數位族語資料庫」就是各種族語資料 (包含書面、手稿、錄音帶、錄影帶、音檔、影音資料等) 經數位化後以電子資料形式儲存的空間 (如下圖所示)，除了典藏族語資料，也讓使用者可以進行檢索，並在合理的規範下，進一步將這些資料做其他教學、研究的用途。

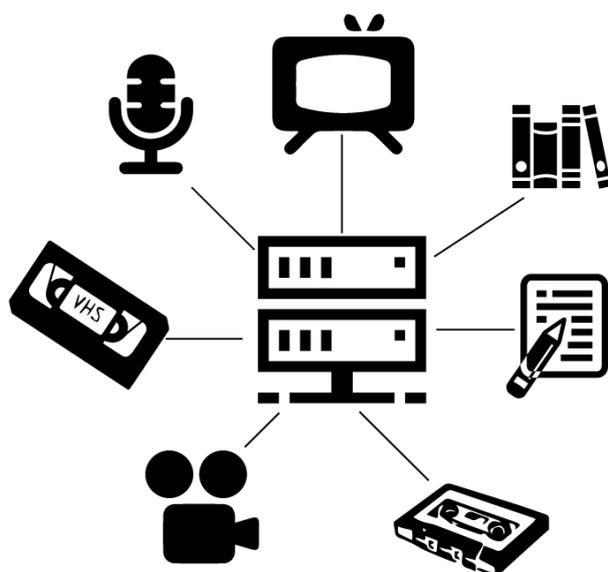


圖 1：資料庫的想像示意圖

當然，基於語料的保存及應用來說，進一步進行語料分析研究的工作也是十分重要的，雖然目前以我們之能還沒辦法做到真正嚴謹、有分析和標記的語料庫形式，但也希望能挑選部分的口語長篇語料或族語新聞先轉譯成族語文字，並逐詞翻譯，這個部分的工作我們也積極規劃及洽談中。

而瞭解我們對數位族語資料庫的想像之後，接下來我們從國內外相關，也一起介紹國內外的相關語言典藏／語料庫網站，做為本計畫參考的基礎。

二、現況分析：參考國內外語言資料庫與語言典藏相關網站

為了瞭解國內外建置相關語言資料的網站，我們瀏覽了很多頁面，但因為語言性質或是所需功能不太相似，就不特別分析，下面將以表格的方式，呈現重要參考網站的歸納與分析，這些網站對於我們規劃系統及其功能，都有莫大的助益。

表 1：

序	名稱（按筆劃）	形式	內容分析
1	AILLA（The	大型	➤ 支援：由澳洲爾本大學 Rachel Nordlinger 教

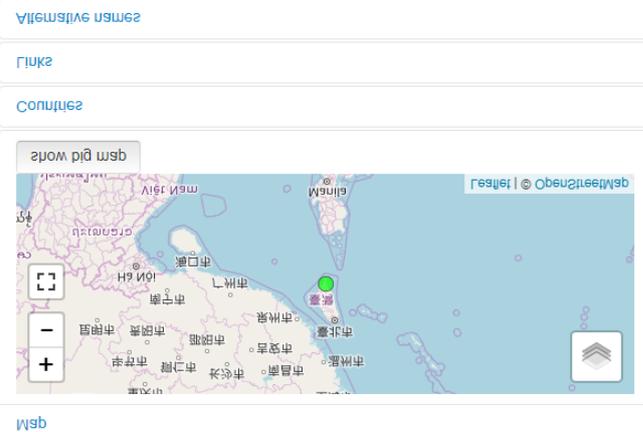
序	名稱 (按筆劃)	形式	內容分析
	Archive of the Indigenous Languages of Latin America), 拉丁美洲原住民族語言檔案館 https://ailla.utexas.org/	綜合資料庫	<p>授及阿德萊德大學 Ian Green 博士合作，並得到 ARC 卓越語言動力中心的資助。</p> <ul style="list-style-type: none"> ➤ 目的：AILLA 的使命是保存拉丁美洲原住民族語言的錄音、文本和其他多媒體的數位語言檔案，並將他們提供給原住民、研究人員和其他有興趣的朋友。 ➤ 特色與功能： <ol style="list-style-type: none"> 1. 網站必須註冊及通過審核。 2. 資料有分級限制，保存的語言資料是有分析過的也有原始資料，他們很歡迎使用者在合理的規範下使用，並協助轉譯成文字或進一步的語言分析。 3. 主要收藏是錄音、影音視頻，內容包含敘事、詠唱、演講、對話、歌曲。 4. 另外也有很多語言文本資料，例如語法的分析、字典或民族誌的現場筆記。 ➤ 分析：這個網站的相關權利義務很清楚，資料也有分級，也在乎族人的隱私及知情同意權。智慧財產權制度都可以做為參考。 <p>檢索頁面</p> 
2	Daly Languages (AUSTRALI	小型語料庫	<ul style="list-style-type: none"> ➤ 支援：由澳洲爾本大學 Rachel Nordlinger 教授及阿德萊德大學 Ian Green 博士合作，並得到 ARC 卓越語言動力中心¹的資助。

¹ <http://www.dynamicsoflanguage.edu.au/>

序	名稱 (按筆劃)	形式	內容分析
	<p>A), 澳洲 Daly 語語料庫</p> <p>http://www.dalylanguages.org/</p>		<p>➤ 目的：針對研究人員和社區提供 Ian Greens 在 1980 年至 1996 年間於澳大利亞北部戴利地區的許多語言的廣泛的實地記錄，記錄和分析。另外還有一個目的是提供 Daly 達利語言的入門資訊以及相關的可用資訊。</p> <p>➤ 特色與功能：</p> <ol style="list-style-type: none"> 1. 需有 PARADISEC 帳號。 2. 語料庫包含大量的田野筆記，以及字典材料，未出版的手稿，其中包含大部分語言的語法描述和語言分析，以及歷史重建。 3. 11 種語言共 157 小時的錄音，部分包含語法、語音、詞彙等文字。所有錄音已經數字化，並在 PARADISEC 中提供。 4. 大量的現場筆記本將掃描後增加至 PARADISEC 中。 <p>➤ 分析：網站資訊清楚，是有目的性的將現有的語料數位化，但資料是儲存在外部，即 PARADISEC 網站。</p> <p>存有之資料</p> <ul style="list-style-type: none"> • Marrithyel, 32 hours, including texts and elicitation (grammatical, phonological, lexical) • Marri Tjevin, 6 hours, grammatical elicitation • Magati Ke, 3 hours, grammatical elicitation • Marri Ngarr, 36 hours, elicitation (grammatical, phonological, lexical) • Marramaninjsji, 6 hours, elicitation (grammatical, phonological, lexical), some texts • Batjamalh, 2 hours, elicitation • Merranunggu, 10 hours, including texts and elicitation (grammatical, phonological, lexical) • Menthe, 7 hours, elicitation • Emmi, 2 hours, elicitation • MalakMalak, 36 hours, elicitation (grammatical, phonological, lexical) • Matngele, 17 hours, elicitation (grammatical, phonological, lexical) <p>語言介紹</p>

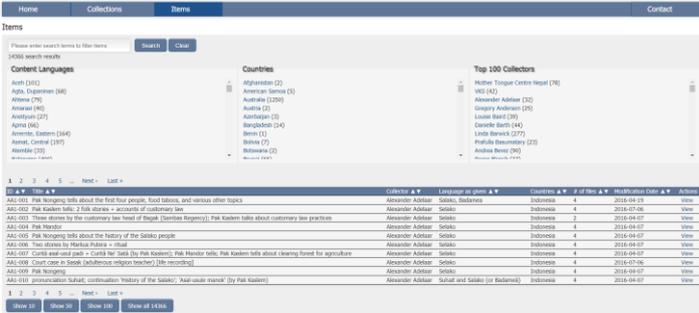
序	名稱（按筆劃）	形式	內容分析
			<p>Marri Tjevin</p> <p>also called: Tanggural, Ngangiyaben, Murintjaran, Murindjabin dialects with: Marrithiyel, Marri Ammu, Marri Dan various spellings: [3]</p> <p>WESTERN DALY GROUPING [3]</p> <p>Marri Tjevin and Marri Ammu are two closely related dialects whose country lies along the coast north of the mouth of the Moyle River, extending up towards the Dashwood Plains. Marri Tjevin occupies the southern part of this area, with Magali Ke and Marri Ngarr to the south. Marri Ammu then has Menthe and Merranunggu to its north and north-east respectively. Their sister dialect Marrithiyel lies to their east. There are a number of active outstations on Marri Tjevin and Marri Ammu country proper, with Marri Tjevin and Marri Ammu communities also resident at Wadeye.</p> <p>While neither Marri Tjevin nor Marri Ammu have been extensively documented, there is general agreement that there are only minor linguistic differences between them, with Tryon (1974) suggesting that their grammars are effectively identical and that they are only distinguished through vocabulary. The community language experts that we have worked with have indicated that there may also be minor phonological differences. Tryon (1968, 1970, 1974) and Kofod (1986a, 1986b), together with sections of Ford (2007), are the only published sources on the dialects, although publications on plant and animal terminology (Northern Territory Government), and a learner's guide (Batchelor College), appear to be under development.</p> <p>It had been thought that no full speakers of either dialect remained. However, two senior Marri Tjevin men, Atu Frank Dumoo (sadly now deceased) and Malbiyin Claver Dumoo, began work in Wadeye with Dr John Mansfield (University of Melbourne) in the development of an extensive set of audio and visual recordings providing detailed documentation of Marri Tjevin language and material culture. The products of this work can be found under the Recordings and Field Notes tab, with additional work being added as it becomes available.</p> <p>Ian Green's corpus of Marri Tjevin consists of 6 hours of recordings of lexical, and grammatical elicitation. The first recording has some textual material in Marri Tjevin (Ambrose Pliatum), Marrithiyel (Bill Parry) and Marri Ngarr (Jimmy Nambatu). The other recordings were made with the following speakers:</p> <ul style="list-style-type: none"> • Les Kuntjul • Colin Worumbu • Edward Nimarluk <p>檔案內容，需連結到 PARADISEC 才能下載</p> <p>Recordings and Field Notes</p> <p>These media files are stored externally at paradisec.org.au. To view them, you need to either sign in or create a free account. Hide this message.</p> <p>John Mansfield (collector), 2012, Marri Tjevin (JM5), Digital collection managed by PARADISEC. [Other Access] DOI: 10.4225/72/5798DA8836497</p> <p>G2-001, Marri Tjevin/Marrithiyel, audio (as spoken by: Nambatu, Parry, Pliatum)</p> <p>G2-002, Marri Tjevin, 10/11/93, audio (as spoken by: Worumbu, Kuntjul)</p> <p>G2-003, Marri Tjevin, 10/11/93, audio (as spoken by: Worumbu, Kuntjul)</p> <p>G2-004, Marri Tjevin, 10/11/93, audio (as spoken by: Worumbu, Kuntjul)</p> <p>G2-005, Marri Tjevin, 13/11/93, audio (as spoken by: Worumbu, Nimarluk)</p> <p>G2-006, Marri Tjevin, 13/11/93, audio (as spoken by: Worumbu, Nimarluk)</p> <p>John Mansfield (collector), 2012, Marri Tjevin (JM5), Digital collection managed by PARADISEC. [Other Access] DOI: 10.4225/72/5798DA8836497</p> <p>Key Published Resources</p> <p>Learning Resources</p> <p>Other Resources</p>
3	<p>Glottolog，世界語言參考文獻網</p> <p>http://glottolog.org/</p>	<p>大型資料庫</p>	<p>➤ 此為協作性的工作，哈拉爾·哈馬斯特羅姆（HaraldHammarström）收集許多書目變成主書目，以及關於世界各種語言證明的家譜關係的大量訊息。</p> <p>➤ 目的：Glottolog 旨在提供一個語言學家需要能夠識別的 languoids（家庭，語言，方言）的綜合列表。每個 languoid 都有一個獨特而持久的名為 Glottocode 的標識符，由四個字母和四個數字組成[abcd1234]。並組織大型書目導入，是一個很大的書目、語言分類查詢網站。</p> <p>➤ 特色與功能：</p> <ol style="list-style-type: none"> 1. 識別語言分類，至 2017 年 11 月 22 日止，有 7384 種口語語言（指第一語言的

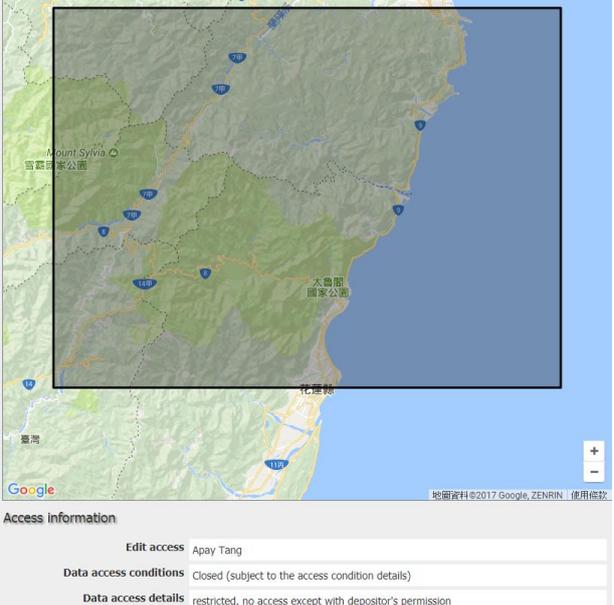
序	名稱 (按筆劃)	形式	內容分析																																																																													
			<p>口語)。</p> <ol style="list-style-type: none"> 2. 搜尋結果會以樹狀結構產生，點擊語別下面會出現相關的書目。 3. 書目有 citation 功能可直接引用。 4. 收集世界語言多樣性遺產。 <p>➤ 分析：這個網站主要是分類語言及語言的關係，再導入相關的書目，並結合地理資訊系統。</p> <p>分類方式</p> <p>Language: Paiwan 📌 ✎</p> <div data-bbox="810 882 1484 1377"> <p>Classification 📌</p> <ul style="list-style-type: none"> - ▼ Austronesian (1275) <ul style="list-style-type: none"> ▶ Atayalic (2) ▶ Bunun ▶ East Formosan (6) ▶ Malayo-Polynesian (1255) ▶ Northwest Formosan (2) ▶ Paiwan ● ▶ Puyuma ▶ Rukai ▶ Tsouic (3) ▶ Western Plains Austronesian (3) <p>Comments on subclassification</p> <p>Robert A. Blust and Victoria Chen 2017 Robert A. Blust 2009</p> </div> <p>書目呈現</p> <table border="1"> <thead> <tr> <th>Details</th> <th>Name</th> <th>Title</th> <th>Any Field</th> <th>id</th> <th>Year</th> <th>Pages</th> <th>Language</th> <th>id</th> <th>Provider</th> <th>id</th> </tr> </thead> <tbody> <tr> <td>🔍</td> <td>Search</td> <td>Search</td> <td>👍</td> <td>SEARCH</td> <td>SEARCH</td> <td>SEARCH</td> <td>SEARCH</td> <td>SEARCH</td> <td>SEARCH</td> <td>SEARCH</td> </tr> <tr> <td>📖</td> <td>Davidson, James O'Rourke 1900</td> <td>The island of Formosa, past and present: history, people, resources, and commercial prospects. The complete, single page plan and outline accompanied by maps and other illustrations</td> <td>✓</td> <td></td> <td>1900</td> <td>500</td> <td>ethnographic, historical, scientific</td> <td>id</td> <td></td> <td></td> </tr> <tr> <td>📖</td> <td>Chiu, Hsueh 2002</td> <td>Paiwan Introduction: "Paiwan Country", Dutch-French</td> <td>✓</td> <td></td> <td>2002</td> <td>400</td> <td>historical</td> <td>id</td> <td></td> <td></td> </tr> <tr> <td>📖</td> <td>Daly, Robert and Williams, Jane 2002</td> <td>One hundred Paiwan tales</td> <td>✓</td> <td></td> <td>2002</td> <td>500</td> <td>id</td> <td>id</td> <td></td> <td></td> </tr> <tr> <td>📖</td> <td>French, George 1902</td> <td>Paiwan dictionary</td> <td>✓</td> <td></td> <td>1902</td> <td>500</td> <td>id</td> <td>id</td> <td></td> <td></td> </tr> <tr> <td>📖</td> <td>Chang, Ann-Hsiuehshuan 2006</td> <td>A Reference Grammar of Paiwan</td> <td>✓</td> <td></td> <td>2006</td> <td>400</td> <td>id</td> <td>id</td> <td></td> <td></td> </tr> </tbody> </table> <p>地理資訊</p>	Details	Name	Title	Any Field	id	Year	Pages	Language	id	Provider	id	🔍	Search	Search	👍	SEARCH	📖	Davidson, James O'Rourke 1900	The island of Formosa, past and present: history, people, resources, and commercial prospects. The complete, single page plan and outline accompanied by maps and other illustrations	✓		1900	500	ethnographic, historical, scientific	id			📖	Chiu, Hsueh 2002	Paiwan Introduction: "Paiwan Country", Dutch-French	✓		2002	400	historical	id			📖	Daly, Robert and Williams, Jane 2002	One hundred Paiwan tales	✓		2002	500	id	id			📖	French, George 1902	Paiwan dictionary	✓		1902	500	id	id			📖	Chang, Ann-Hsiuehshuan 2006	A Reference Grammar of Paiwan	✓		2006	400	id	id								
Details	Name	Title	Any Field	id	Year	Pages	Language	id	Provider	id																																																																						
🔍	Search	Search	👍	SEARCH	SEARCH	SEARCH	SEARCH	SEARCH	SEARCH	SEARCH																																																																						
📖	Davidson, James O'Rourke 1900	The island of Formosa, past and present: history, people, resources, and commercial prospects. The complete, single page plan and outline accompanied by maps and other illustrations	✓		1900	500	ethnographic, historical, scientific	id																																																																								
📖	Chiu, Hsueh 2002	Paiwan Introduction: "Paiwan Country", Dutch-French	✓		2002	400	historical	id																																																																								
📖	Daly, Robert and Williams, Jane 2002	One hundred Paiwan tales	✓		2002	500	id	id																																																																								
📖	French, George 1902	Paiwan dictionary	✓		1902	500	id	id																																																																								
📖	Chang, Ann-Hsiuehshuan 2006	A Reference Grammar of Paiwan	✓		2006	400	id	id																																																																								

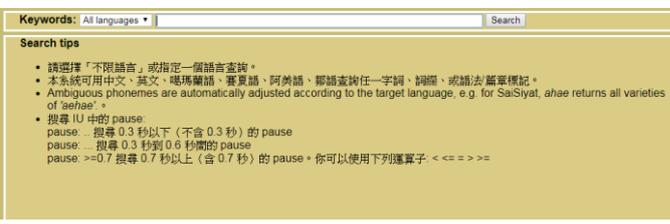
序	名稱（按筆劃）	形式	內容分析
			
4	<p>PARADISEC (Pacific and Regional Archive for Digital Sources in Endangered Cultures)，太平洋地區檔案數位資源瀕危文化庫</p> <p>http://www.paradisec.org.au/</p>	大型資料庫	<ul style="list-style-type: none"> ➤ 支援：由澳洲雪梨大學、墨爾本大學、澳洲國立大學三所大學組成的，接受捐款。 ➤ 目的：將世界上小語種的語言以數位方式保存。 ➤ 特色與功能： <ol style="list-style-type: none"> 1. 目前超過 1000 種語言。 2. 主要專注於數位化錄音磁帶，但另外開放可以存放音檔衍生而成的文本材料（論文、字典、文章等）。 3. 需註冊。 4. 提供語言與音樂數據管理，錄音及數據連接的培訓。 5. 檢索功能可點地圖，或從搜尋欄打關鍵字，或從語言別、國家、前 100 的收藏或是可以自己創建目錄。 6. 結合地理資訊系統。 7. 標準化檔案，亦可在 OLAC（開放語言檔案社群網）中搜尋資料。 8. 與當地機構合作（如所羅門群島博物館）提供備份和數據保存服務。 9. 付費磁帶修復及數位化²。 10. 有各式協助指南³（歸檔工作、數據儲存、入門介紹、Metadata 說明等）。

² <http://www.paradisec.org.au/deposit/costing-a-deposit/>

³ <http://www.paradisec.org.au/resources/downloads/>

序	名稱 (按筆劃)	形式	內容分析																																					
			<p>➤ 分析：網站資訊清楚，資料庫訊息一目瞭然，結合地理資訊系統，檢索方式多元也很方便。</p> <p>檢索頁面</p>  <p>資料明細</p> <p>Collection details</p> <table border="1"> <tr><td>Collection ID</td><td>AIT1</td></tr> <tr><td>Title</td><td>Recordings of Taroko (Taiwan)</td></tr> <tr><td>Description</td><td>Recordings of narratives in Taroko (Taiwan).</td></tr> <tr><td>Archive link</td><td>http://catalog.paradisec.org.au/repository/AIT1</td></tr> <tr><td>Collector</td><td>Apay Tang Find similar</td></tr> <tr><td>Operator</td><td></td></tr> <tr><td>Originating university</td><td>University of Hawaii at Manoa</td></tr> <tr><td>Countries</td><td>Taiwan - TW To view related information on a country, click its name</td></tr> <tr><td>Languages</td><td>Taroko - trv To view related information on a language, click its name</td></tr> <tr><td>Region / village</td><td></td></tr> <tr><td>Cite as</td><td>Apay Tang (collector), 1997; <i>Recordings of Taroko (Taiwan)</i> (AIT1), Digital collection managed by PARADISEC. [Closed Access] DOI: 10.4225/72/56E7A74E33FB7</td></tr> </table> <p>所存資料 (圖為未開放)</p> <p>Items in Collection (4)</p> <table border="1"> <thead> <tr> <th>Item</th> <th>Title</th> <th>Actions</th> </tr> </thead> <tbody> <tr> <td>001 closed</td> <td>traditional Truku stories</td> <td>View</td> </tr> <tr> <td>002 closed</td> <td>daily stories</td> <td>View</td> </tr> <tr> <td>003 closed</td> <td>Daily and traditional stories</td> <td>View</td> </tr> <tr> <td>004 closed</td> <td>Tama visitation</td> <td>View</td> </tr> </tbody> </table> <p>地理資訊</p>	Collection ID	AIT1	Title	Recordings of Taroko (Taiwan)	Description	Recordings of narratives in Taroko (Taiwan).	Archive link	http://catalog.paradisec.org.au/repository/AIT1	Collector	Apay Tang Find similar	Operator		Originating university	University of Hawaii at Manoa	Countries	Taiwan - TW To view related information on a country, click its name	Languages	Taroko - trv To view related information on a language, click its name	Region / village		Cite as	Apay Tang (collector), 1997; <i>Recordings of Taroko (Taiwan)</i> (AIT1), Digital collection managed by PARADISEC. [Closed Access] DOI: 10.4225/72/56E7A74E33FB7	Item	Title	Actions	001 closed	traditional Truku stories	View	002 closed	daily stories	View	003 closed	Daily and traditional stories	View	004 closed	Tama visitation	View
Collection ID	AIT1																																							
Title	Recordings of Taroko (Taiwan)																																							
Description	Recordings of narratives in Taroko (Taiwan).																																							
Archive link	http://catalog.paradisec.org.au/repository/AIT1																																							
Collector	Apay Tang Find similar																																							
Operator																																								
Originating university	University of Hawaii at Manoa																																							
Countries	Taiwan - TW To view related information on a country, click its name																																							
Languages	Taroko - trv To view related information on a language, click its name																																							
Region / village																																								
Cite as	Apay Tang (collector), 1997; <i>Recordings of Taroko (Taiwan)</i> (AIT1), Digital collection managed by PARADISEC. [Closed Access] DOI: 10.4225/72/56E7A74E33FB7																																							
Item	Title	Actions																																						
001 closed	traditional Truku stories	View																																						
002 closed	daily stories	View																																						
003 closed	Daily and traditional stories	View																																						
004 closed	Tama visitation	View																																						

序	名稱 (按筆劃)	形式	內容分析
			
5	<p>臺大臺灣南島語多媒體語料庫</p> <p>http://corpus.linguistics.ntu.edu.tw/index_zh.php。</p>	<p>中型語料庫</p>	<ul style="list-style-type: none"> ➤ 支援：由臺灣大學語言學研究所建立，原為國立臺灣大學資訊電子科技整合研究中心「多媒體整合實驗室」子計畫之一(2001-2003)，由臺灣大學語言學研究所黃宣範、蘇以文及宋麗梅教授共同主持。 ➤ 目的：以臺灣南島語為主軸，藉由資訊科技運用，建構起語言資料典藏及後設資料之標準，有系統彙整、轉寫、保存彌足珍貴的口述語料。 ➤ 特色與功能： <ol style="list-style-type: none"> 1. 語料庫中建置賽夏語、噶瑪蘭語、鄒語及阿美語資料庫。有聲音或影像檔。 2. 未來展望，希望為臺灣南島語建置語言資料典藏及後設資料之標準。並以半自動化為目標。 3. 不論是語言學者或是一般大眾，只要是對南島文化及語言有興趣者，藉由本語料庫，即可應用最便捷、多元的介面，取得豐富而珍貴的語料。 4. 語料庫的基本原則，普及性、口語性、一致性、跨語言性、交叉性、相容性、便利性、多元性。 5. 語料文本基本上有 Pear Stories 和 Frog

序	名稱 (按筆劃)	形式	內容分析
			<p>Story，故有範圍限制，較能夠做跨語言的比較分析。</p> <p>6. 語料呈現方式為族語句子、註解（英）、註解（華）、完整翻譯（英）、完整翻譯（華），有族語、華語、英語翻譯，讓國際人士都方便查詢。</p> <p>➤ 分析：網站資訊清楚，資料庫訊息一目瞭然。</p> <p>依語別檢索</p>  <p>關鍵字搜尋 (可跨語言)</p>  <p>語料呈現方式</p> 
6	<p>臺灣南島語數位典藏 (目前已停用)</p> <p>http://languagearchives.sinica.edu.tw/cht/index</p>	<p>中型語料庫</p>	<p>以下內容整理自齊莉莎、余清華 (2007)</p> <p>➤ 支援：是國科會 NDAP 機構計畫下分項主題「語言典藏」的子計畫之一，整體建置作業之規劃暨主持人為齊莉莎女士，計畫分二期 2002-2006、2007-2011，現已停用。</p> <p>➤ 目的：建立一數位圖書館(digital library)，期</p>

序	名稱（按筆劃）	形式	內容分析
	php.html		<p>能收集、保存、編輯及透過網路散播語言資源，供使用者存取已錄製及記音的南島語。</p> <p>➤ 特色與功能：</p> <ol style="list-style-type: none"> 1. 語料庫中經語料分析的有魯凱語、賽夏語、泰雅語、布農語、排灣語、卑南語。其他語料輸入有布農語、鄒語、巴宰語、卡那卡那富語、西拉雅語、魯凱語、排灣語、阿美語。 2. 另外也建置語言學相關書目的資料庫。 3. 有語言地理資訊系統了解，臺灣南島語的分布圖，並觀察同源詞與非同源詞的分布情形。 4. 有進階檢索功能，也提供詞彙用字索引，讓使用者檢視詞彙的分布。 5. 提供原始語料與重新分析語料做對照。 6. 有特殊符號代碼，方便查詢時使用。 7. 因各族書寫不一致，故使用 IPA 記音，也較不容易混淆。 8. Metadata 包含語料的側寫（如：題目、語言及方言、文體），田野調查（如：發音人、田野工作者/謄寫者/編輯者/翻譯者、資料收集/編輯/翻譯的日期）以及管理策略（版權宣告）。 9. 開發 AnnoTool（提供分析用縮寫，華語英語縮寫可以翻譯）及 Chkgloos（檢，查不同語言間的詞彙數量）兩種軟體，協助分析者能提供更正確及一致性高的語料。 <p>➤ 分析：網站功能很完整，語料資源豐富，又結合地理資訊系統及建置書目，是確實可以產出學術研究的多功能語料庫。但很可惜現在計畫沒有繼續因此停用中。</p> <p>檢索頁面</p>

序	名稱 (按筆劃)	形式	內容分析
			 <p>The screenshot displays the Academia Sinica Formosan Language Archive website. At the top, there is a navigation bar with the site's name and logo. Below this, a search interface is visible, including a search box and several filter options: '設定語別搜尋範圍' (Set language search range), '設定文體搜尋範圍' (Set text style search range), '設定語料呈現方式' (Set corpus display method), and '設定語料尋找方式' (Set corpus search method). The '執行' (Execute) button is highlighted. Below the search interface, a list of linguistic data entries is shown, each with a unique ID (e.g., 01.001.a) and a corresponding audio player. The entries include transcriptions in Romanized Formosan and their corresponding Chinese translations.</p>

序	名稱（按筆劃）	形式	內容分析																																										
			<p>01.001.b  日本人來了以後就（開始）學我們萬山和那東斯（排灣及魯凱族）的話。 0naŋi a-kaavaŋi-ŋa-ða ʒaomo mani a a ʒitsaʒotsaʒo ʒir 那 分句名物化-動態-非限定 亦-已經-他 屬格 日本人 敬 動態-非限定 亦 動態-處與式 畢 羅</p> <p>01.001.c  他們比較這兩種語言後，（就發現）我們的話有許多（和）魯凱族是一樣的。 mani poa a-pa-rama-ramao kapa-ka-ðao-ðao ta-ʒini ao 敬 動態-非限定 亦 動態名物化-動態-非限定 相同-重疊-相同 * 狀態-非限定-重疊-多 主語名物化-動</p> <p>段落顯示</p> <p>1.  onaŋi ʒaamaðalaa-nai ta-piʒa-aa-na-ða po-a atsa ʒoponoho m-ia. 0naŋi a-kaavaŋi-ŋa-ða ʒaomo mani a a ʒitsaʒotsaʒo ʒina vaha-nai ʒoponoho la tali- ao- aobo. mani poa a-pa-rama-ramao kapa-ka- ðao-ðao ta-ʒini ao ʒi tali- akai ʒina vaha-nai. mani a -inamo patoʒo: "maʒ-itolo-tolo o-nomi" m-ia. mani jhoʒo-nai ʒina ʒoponoho "ka- ʒa- ta aanai" m-ia-nai. 0naŋi a-ki-kaavaŋi-na-ða ʒaomo, o- jhoʒo-ka-nai maʒ-itolo-tolo o-mita ɖakai m-ia. kasa-ta ʒoponoho vaha-ta om-ia- nai.</p> <p>我們的祖先萬山自稱是萬山人。日本人來了以後就（開始）學我們萬山和那東斯（排灣及魯凱族）的話。他們比較這兩種語言後，（就發現）我們的話有許多（和）魯凱族是一樣的。於此他會告訴我們說：「你是阿宗（的）」。我們萬山人（才）知道「原來我們是（屬於魯凱族）」。日本人還沒來之前，我們不知道我們和魯凱族同宗。我們只（認為）咱們的話（我）是萬山話。</p> <p>內容檢索</p> <p>內容檢索</p> <p> 回上頁  使用說明  回首頁</p> <p>■搜尋範圍－語言別：魯凱語/萬山 文體：全部 語料搜尋：搜尋模式</p> <div data-bbox="831 1003 1469 1234"> <p>關鍵詞示例</p> <p>1. 關鍵詞以原文輸入時,若使用到 IPA 音標,請點選對應的代碼如下:</p> <table border="1"> <tr> <td>IPA</td> <td>ð</td> <td>ŋ</td> <td>ɖ</td> <td>a</td> <td>l</td> <td>ʔ</td> <td>i</td> <td>θ</td> <td>ʌ</td> <td>ɣ</td> </tr> <tr> <td>代碼</td> <td>240</td> <td>331</td> <td>598</td> <td>601</td> <td>621</td> <td>660</td> <td>616</td> <td>952</td> <td>654</td> <td>620</td> </tr> </table> <table border="1"> <tr> <td>IPA</td> <td>o</td> <td>ʃ</td> <td>æ</td> <td>œ</td> <td>β</td> <td>ʍ</td> <td>ʎ</td> <td>ʒ</td> <td>ʈ</td> </tr> <tr> <td>代碼</td> <td>216</td> <td>643</td> <td>230</td> <td>339</td> <td>946</td> <td>649</td> <td>404</td> <td>607</td> <td>648</td> </tr> </table> <p>2. 關鍵詞以註解輸入時,請輸入可能出現的字(如:酒、芒草、書)</p> <p> 執行  清除</p> </div>	IPA	ð	ŋ	ɖ	a	l	ʔ	i	θ	ʌ	ɣ	代碼	240	331	598	601	621	660	616	952	654	620	IPA	o	ʃ	æ	œ	β	ʍ	ʎ	ʒ	ʈ	代碼	216	643	230	339	946	649	404	607	648
IPA	ð	ŋ	ɖ	a	l	ʔ	i	θ	ʌ	ɣ																																			
代碼	240	331	598	601	621	660	616	952	654	620																																			
IPA	o	ʃ	æ	œ	β	ʍ	ʎ	ʒ	ʈ																																				
代碼	216	643	230	339	946	649	404	607	648																																				

序	名稱 (按筆劃)	形式	內容分析																																
			<p>線上學習網，是研究和學習達悟語很好的工具。</p> <p>語料分類</p> <div data-bbox="794 465 1487 719"> <p style="text-align: center;">[達悟語料]</p> <p style="text-align: center;">故事篇.生活篇.祭典與傳統文化篇.政見發表篇.女性議題篇.動畫篇.其他篇</p> <table border="1"> <thead> <tr> <th>主分類</th> <th>次分類</th> <th>篇序</th> <th>達悟篇名</th> <th>中文篇名</th> </tr> </thead> <tbody> <tr> <td></td> <td>意外經驗</td> <td>01</td> <td>micahas a longtsang si Kawcincong</td> <td>越獄犯高連忠</td> </tr> <tr> <td></td> <td>學校生活</td> <td>02</td> <td>mapakazakazakat a yalalaw so sinsi</td> <td>裝死嚇老師</td> </tr> <tr> <td></td> <td rowspan="4">部落生活</td> <td>03</td> <td>aney o hango ko!</td> <td>還我林投果!</td> </tr> <tr> <td></td> <td>04</td> <td>man'acing so kols</td> <td>圍豬腳</td> </tr> <tr> <td></td> <td>05</td> <td>miarap</td> <td>打群架</td> </tr> <tr> <td></td> <td>06</td> <td>teymangangasien ko ya</td> <td>我也是個有愛心的人</td> </tr> </tbody> </table> </div> <p>語料分析</p> <div data-bbox="794 763 1481 1137"> <p>1. yaken rana am, mina apia dang o kakawan a.</p> <p>我主格 已經 呢 曾經 很好 當時 主格 天氣 助</p> <p>1.S.NOM already PAR PA good then NOM weather PAR</p> <p>k-aro namem a ka-k-anak-an</p> <p>故-很多 我們屬格除 緊 疊-名綴-小孩子-名綴</p> <p>CON-many 1.P.GEN.EXCL LIN RED-NF-child-NF</p> <p>a mi-al-alam do Jicinikcikan ito a.</p> <p>緊 主格-疊-玩耍 處格 地名 那個 助</p> <p>LIN AF-RED-play LOC PLN that PAR</p> <p>【我呢，那天天氣很好，所以我們很多小孩子就都去Jicinikcikan遊玩】</p> <p>As for me, one day the weather was very good, and a lot of us children went to Jicinikcikan to play.</p> </div> <p>關鍵字查詢</p> <div data-bbox="794 1193 1481 1608"> <p style="text-align: center;">關鍵字查詢</p> <div style="border: 1px solid black; padding: 5px;"> <p style="text-align: center;">查詢方式</p> <p> <input checked="" type="radio"/> 達悟語查詢 <input type="radio"/> 中文查詢 <input type="radio"/> 英文查詢 </p> <p style="text-align: center;">查詢來源</p> <p> <input checked="" type="radio"/> 所有來源 <input type="radio"/> 四冊教材 <input type="radio"/> 二十篇語料 <input type="radio"/> 聖經 </p> <p style="text-align: center;">查詢種類</p> <p> <input checked="" type="radio"/> 部份查詢 <input type="radio"/> 全字查詢 </p> <p style="text-align: center;">請輸入關鍵字:</p> <div style="border: 1px solid black; height: 20px; width: 100%;"></div> <p style="text-align: center;">開始查詢</p> </div> </div>	主分類	次分類	篇序	達悟篇名	中文篇名		意外經驗	01	micahas a longtsang si Kawcincong	越獄犯高連忠		學校生活	02	mapakazakazakat a yalalaw so sinsi	裝死嚇老師		部落生活	03	aney o hango ko!	還我林投果!		04	man'acing so kols	圍豬腳		05	miarap	打群架		06	teymangangasien ko ya	我也是個有愛心的人
主分類	次分類	篇序	達悟篇名	中文篇名																															
	意外經驗	01	micahas a longtsang si Kawcincong	越獄犯高連忠																															
	學校生活	02	mapakazakazakat a yalalaw so sinsi	裝死嚇老師																															
	部落生活	03	aney o hango ko!	還我林投果!																															
		04	man'acing so kols	圍豬腳																															
		05	miarap	打群架																															
		06	teymangangasien ko ya	我也是個有愛心的人																															

參、數位族語資料庫計畫執行內容

根據前一章的介紹，我們初步瞭解了數位族語資料庫的概念，以及認識國內外語言典藏網站的功能，而接下來要進入實際的討論，並回顧我們今年執行計畫的工作內容。

一、數位族語資料庫建置目的

建置數位族語資料庫的核心目的，就是要進行語言資料的保存及應用。2017年6月14日公布的原住民族語言發展法第九條亦明訂，需建置原住民族語言資料庫，積極保存原住民族語料。

因此，本計畫第一階段的目的，即是建置「數位族語資料庫」，內含語料典藏、族語新聞轉譯及原住民族語言研書目等，未來可供民眾檢索使用。

二、計畫執行方式

本計畫今年主要的工作目標，即是建置數位族語資料庫，除了將搜集語料數位化，也檢視2017年的數位族語資料庫語料清單，將部分可使用的語料試建置於資料庫中。而系統初步建置後，我們開始利用逐字稿系統及新聞編輯系統進行族語新聞的轉譯編輯工作。另一方面本年度共辦理9場研商會議，包含建置工作、轉譯工作及對系統修正的討論等，如下圖：

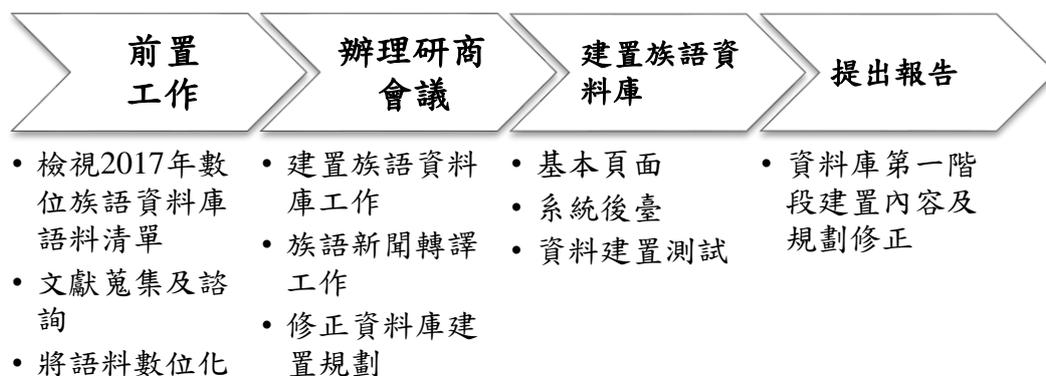


圖 2：數位族語資料庫—研究執行架構

以下我們就分項說明計畫執行的方式：

(一) 族語資料及文獻調查

為將珍貴的族語文獻資料彙集並進行數位典藏，以利族語資料之保存、研究及應用，因此 2017 年我們要透過國家圖書館「臺灣書目整合查詢系統」、「臺灣原住民族圖書資訊中心」、「臺灣原住民族數位典藏資料庫」等系統性檢索，並以及田野調查的方式，初步彙整相關的語料清單⁴。

我們將語料清單分為書面語料、口語語料及其他影音語料，根據語料清單調查的結果，教材的數量最多有 375 筆，而詞典類也有 51 筆，其餘類型如文學創作、口傳文學等也不多。依族群來看，阿美族有 109 筆是數量較多，卡那卡那富族及拉阿魯哇族則是僅列 6 筆，也反應了多方言、單方言及族群人數多寡之差距，因此在未來資料的建置時程上，我們必須加快腳步，也一起配合搶救瀕危語言計畫，盡量保存我們的語言資料。

經過 2017 年田調訪談，計畫面臨到一個困境，即無法於現階段取得私人族語文獻（語料）持有相關的資料清單，訪談的結果一致都是希望網站能先建立起來，並且確立智慧財產權的保障，讓資料安全且是有限制的被使用，另外也提到若能提供資料檔案的修復，亦是一大吸引。

因此依族語語料清單，今年我們先以 1999 年至 2000 年由原住民族委員會委託學者調查的長篇語料中的太魯閣族語，先做部分的數位化，該語料是由口語轉譯成文字，是十分重要且珍貴的資料。一方面我們也再諮詢當時計畫主持人楊盛涂耆老，他不吝分享，授權提供 40 卷錄音帶，我們目前亦先數位化，有一些部分仍需要修復或技術處理，明年我們會優先處理。有了這樣的基礎資料，可以增加族語資料庫的典藏，以建立外界對於資料庫的信任感，而族人或研究者所持有的資料，就需要未來網站設立並有基礎後再進一步推廣。而未來我們會繼續增加語料，讓資料庫的語言資料更豐富。

⁴數位族語資料庫語料清單見原住民族語言研究發展中心 2017 年「原住民族語言研究發展中心數位族語資料庫建置評析報告」。

(二) 田野訪談

2017年我們針對數位族語資料庫的建置規劃及調查個人所持有的族語文獻清單，訪談諮詢對象共38人次，重點整理如下：

問題類別	建議及討論重點
調查經驗／資料取得／研究論理／授權	<ol style="list-style-type: none"> (1) 需要與族人有良好的互動，另外需遵守研究論理，不該發布的族群重要價值，就不應該發布，才能得到族人的信任可以調查更多資料。 (2) 曾經幫忙族人修復卡帶再數位化，因此取得早期的資料。 (3) 智財權一直是數位化會面臨的問題，要去思考合理的版權費用分配，包含著作人（例如例研人員）、報導人，以及讓著作人選擇開放的程度。重點是要有合理妥善的制度，及安全的管理才會讓人願意將心血公開。 (4) 語言的資料，在跨領域的學科裡面都會有，像是音樂系也會有要分析歌詞，或是錄下聚會吟唱加上談話的自然語音。 (5) 調查完送上一份調查訪問或是拍照的資料，族人也許也會很樂意受訪。
田調資料的保存	<ol style="list-style-type: none"> (1) 早期有卡帶、MD到現在有錄音筆，還有錄影設備，分類建檔，並設置清單，可以學習圖書管理的方式，有標準化的編碼整理資料會很清楚。 (2) 資料的數位化是必然，實體物件可能會因天災人禍消失，但數位資料需注意早期燒錄光碟也有到後期就不能讀取了。 (3) 數位化儲存會需要非常大量的儲存空間，還有備分問題。 (4) 通常有原始調查資料，另外也會將需要的部分分割另存。
族語資料庫的建置	<ol style="list-style-type: none"> (1) 應該要有國家級的單位來統籌，否則很多資料無法跨單位取得，層級太低，規模也做不大。 (2) 資料庫建立 Metadata 需先規範好。 (3) 建置族語資料庫或語料庫需要注意每個語言別的差異。

	(4) 族語文字書寫的一致性也很重要，或是要給予差異的準則。
語料的提供	(1) 建議先完成資料庫之後，再推廣以取得更多資料。 (2) 有全族語的廣播節目，也很適合做為族語資料庫的資料。

而今年雖主要是在建置數位族語資料庫，我們亦繼續蒐集資料，楊盛涂耆老提供 40 卷錄音帶，我們先進行數位化工作，預計 2019 年進行語料修復以及將部分音檔進一步做整理。

另外，我們也嘗試以語言學調查中知名的 pear stories 梨子的故事和 frog story 青蛙的故事來請耆老以講故事的方式來保存語料，未來我們也會持續搜集相關語料。

(三) 研商會議

2017 年度一共辦理八場研商會議，邀請語言學、傳播學、語料庫專業、工程師（詞典數位化、語料庫、語音合成、資料庫）、研究人員、轉譯人員、原文會（族語新聞）等，依場次整理重點如下表：

表 2：研商會議建議及討論重點整理

場次	研商會議建議及討論重點
1	(1) 若要積極建置資料庫，可以分區和各大學校院合作，特別是原住民專班，培訓學生田調的能力，再將資料建置於資料庫，是很好的合作模式。 (2) 國外的語言保存網站模式值得參考，語料文獻持有者可以將檔案資料給予該網站，他們會保存實體物件，並協助數位化，原持有者可以選擇物件公開的程度。 (3) 族語資料庫的經營需要持續的人力及資源維持。 (4) 未來可以發展從各族知識體系來建構的語言學習平臺，讓學習者能多元的學習「語言」，更能學習語言背後的「文化」。
2	(1) 由原民會教文處協助語發中心與原文會的合作討論。 (2) 族語新聞的保存、新聞稿及流程說明。 (3) 與原文會初步討論族語新聞的合作方式，下次會議可實地參訪。

場次	研商會議建議及討論重點
3	(1) 至原文會新聞部實地參訪，含片庫及族語新聞編輯部的狀況及使用的系統介紹。 (2) 取得新聞檔案的方式有三種，需再依需求評估。 (3) 要取得族語新聞授權及 INEWS 系統讀取權限皆需再研商。 (4) 與族語主播合作轉譯新聞事宜，尊重主播個人意願，再議。
4	(1) 萌典 OCR 技術可參考。 (2) 從程式設計的角度，要做的事情、處理的資料、呈現的方式都必須思考清楚，才容易規劃。 (3) 資料典藏可參考文化部「國民記憶庫—臺灣故事島」網站或是臺灣原住民族數位典藏資料庫的形式。 (4) 族語新聞與原文會合作，最好可以開發轉譯標記系統。 (5) 需要很多語料，至少 80 萬以上才適合開發自動轉譯系統。 (6) 後續可以再線上討論，提供系統開發意見，已建立 FB 討論群組。
5	(1) 以語言學家的觀點來看，會比較建議做一個語言，專精、精確就好，但也許與政策有落差，與其放很多不知道正確與否的資料，也許就先做像族語新聞的轉譯，只要先轉譯成文字，即便未分析，有逐詞翻譯就很好用了。跨語言選擇同一篇新聞，也有利於跨語言比較。 (2) 要考慮語言資料庫不要只是「典藏」而是要像活水般流動，另外也要考慮是否能夠永續經營，否則可能會白費工。 (3) 若未來有調查及分析語料，需注意若是以族群的真實文化為主題，報導人提供的故事本身是否為編造或是有謬誤的存在。 (4) 若未來建立學習者語料庫，需要思考「偏誤標記」誰是標準誰是偏誤，要很小心處理。 (5) 可善用 Toolbox、Lexique Pro 等工具。
6	(1) 轉譯及標記工作實際的工作時間，可能是音檔的 30 倍以上，收到音檔最好盡快完成轉譯。 (2) 西拉雅語的復振很大一部分是靠文獻來回復，因此能夠保存大量的語音、文字檔案，確實是很重要的。故資料庫的規劃是可行。
7	(1) 針對目前規劃的系統需求評估及研商。

場次	研商會議建議及討論重點
	(2) 欲建置輸入語料的後臺系統及自訂詞庫、建置語言資料目錄清單系統。 (3) 方向大致可行，細節下次會議提供。
8	(1) 與原文會進行實際的合作溝通，取得合作共識。 (2) 討論包括族語新聞的版權、INEWS 的使用、新聞存取的方式、族語新聞的後續應用、族語轉譯的工作、如何增加族語新詞在族語新聞中的使用頻率。

而今年我們邀請語言學者、資訊工程學者、工程師、族人等進行共 9 場研商會議，會議內容包含資料庫建置工作、族語新聞轉譯工作、資料庫規劃修正等，會議記錄於附錄，各場次討論重點摘要如下：

表 3：研商會議建議及討論重點整理

場次	研商會議建議及討論重點
1	*建置前評估及建議 (1) 基本功能建置，包含資料庫後臺、網站、檢索功能。 (2) 資料庫網站可併入原住民族語言研究書目，以利資料整合。 (3) 建置之資料以多種形式測試。 (4) 業界廠商價格過高，故建議尋找個人工作團隊，以不同專長分工協助資料庫建置，另需注意採用開放程式編輯，且需拿到完整程式碼。 (5) 會後提出資料庫建置需求書。 (6) 逐字稿系統可參考 oTranscribe 等開放程式編寫。
2	*建置前與工程師討論 (1) 本次會議主要與工程師團隊討論規劃的可行性 (2) 基本上如需求書規劃內容執行，細節後續每一階段都再予以調整。 (3) 需進一步提供相關 Metadata 內容。
3-7	*建置工作階段會議 (1) 工程師說明資料庫及網站建置情形。 (2) 進行系統及網站建置之工作協調。
8	*族語新聞逐字稿及轉譯工作 (1) 進行轉譯工作說明
9	*資料庫規劃修正討論

場次	研商會議建議及討論重點
	(1) 針對目前資料庫建置情形給予建議 (2) 未來規劃修正方向

三、數位族語資料庫系統建置內容

今年我們以 2017 年的資料庫建置規劃為基礎，並經過研商會議的討論，我們邀請工程師團隊協助進行資料庫建置。目前第一期的建置內容已完成，「原住民族語言研究發展中心數位族語資料庫」網址為 <http://dailt.ilrdc.tw>。以下將分項介紹目前資料庫系統及網頁的建置情形：

(一) 資料庫基本頁面

1. 頁面規劃及架構：

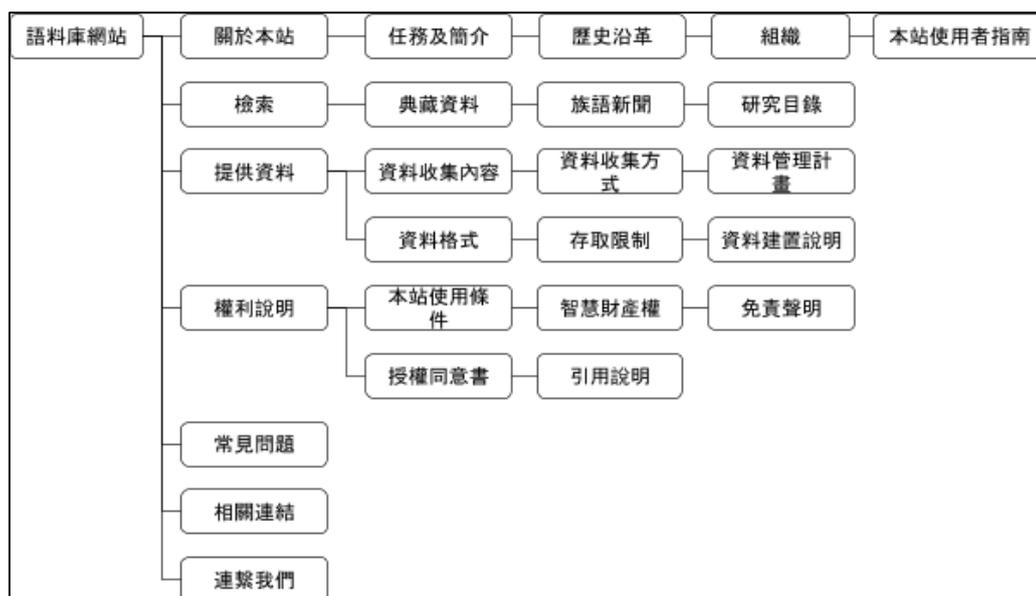


圖 3：數位族語資料庫頁面架構圖

2. 網站基本頁面介紹：

(1) 網站 logo





(2) 上方導覽列

[首頁](#)
[最新消息及公告](#)
[典藏資料檢索](#)
[族語新聞檢索](#)
[臺灣原住民族語言研究書目檢索](#)
[相關連結](#)

■ 首頁

[首頁](#)
[最新消息及公告](#)
[典藏資料檢索](#)
[族語新聞檢索](#)
[臺灣原住民族語言研究書目檢索](#)
[相關連結](#)



DAILT ILRDC
原住民族語言研究發展中心
數位族語資料庫
The Digital Archive of the Indigenous Languages of Taiwan

歡迎來到ILRDC數位族語資料庫

2018年本站開始建置，旨在以數位化方式保存及典藏臺灣原住民族語言資料，因此除了持續收錄新資料(以新增原音)並進行數位化檔案之外，也積極與原文會合作，開始進行部分族語相關新目的族語文字翻譯。另一方面也希望彙整臺灣原住民族語言研究相關書目，提供學界及對原住民族語有興趣之一般民眾、研究人員、族語教學人員等參考，使寶貴的族語語料及研究資料能流通及分享給眾人，也期盼本站所典藏的資料能與原音提供資料者的權益下再應用，以發展更多研究及文本，形成良性的循環。

最新消息

2018.11.16	傳稿
2018.11.16	傳稿
2018.11.09	訪談
2018.10.26	面試 2018
2018.10.26	這是傳稿

[更多消息](#)



有關使用本站重要訊息

1. 註冊 必須同意本站規則，註冊並登入才能觀看、下載及使用資料。
2. 一般使用者 可依所需資料形式進行檢索。
3. 提供資料者 若您想提供您的資料，請點擊連結。



本站收藏

本站收藏的核心為由原住民族語言音檔和視頻，包含故事、演講、對話等，部分資料會轉譯為族語文字。

另一方面也開始轉譯族語新聞，比臺灣原住民族語言研究員目，我們致力於典藏族語資料並將其數位化，希望這些資料能被應用，發展更多族語文本或發展更多研究，若您有任何問題或建議，請聯繫我們。

關於本站

- 任務及簡介
- 歷史沿革
- 組織
- 本站使用說明

檢索

- 典藏資料
- 族語新聞
- 臺灣原住民族語言研究書目

提供資料

- 資料收藏內頁
- 資料提供方式
- 資料管理計畫
- 資料格式
- 存取限制
- 資料權屬說明

權利說明

- 本站使用條件
- 解編計畫
- 免責聲明
- 版權與授權
- 引文規範

常見問題

- 相關連結
- 聯繫我們

原住民族語言研究發展中心
Indigenous Languages Research and Development Center

地址: 10093台北市中正區科發路一段63號 6F-10F
電話: (02)2341-8108
傳真: (02)2341-8256
郵箱: ind63@gmail.com

原住民族語言研究中心
Copyright © 2018 Council of Indigenous Peoples.
All Right Reserved.

■ **最新消息及公告**

消息公告

全部 9 活動 2 典藏資訊 5 研究目錄 2 其他 0

日期	分類	標題	內容
2018-11-16	【活動】	標題	hihihihi....
2018-11-16	【活動】	標題
2018-10-09	【研究目錄】	title	www.google.com url.....
2018-10-09	【研究目錄】	測試 2018	測試 2018 測試 2018...
2018-10-09	【典藏資訊】	這是標題	內容內容www.google.com....
2018-10-05	【典藏資訊】	這裏才是首頁	Tinymce 測試 解壓縮後就可以打開....
2018-10-05	【典藏資訊】	這裏是首頁	這裏是首頁啦 首頁....
2018-10-03	【典藏資訊】	1	1....
2018-09-23	【典藏資訊】	這裏才是首頁	Tinymce 測試 解壓縮後就可以打開....

■ 典藏資料檢索

- 目前能只設定基本檢索關鍵字
- 搜索結果顯示在下方

典藏資料檢索

z

Show 1 to 4 entries

收藏主題(華語)	收藏主題(族語)	族語別	方言別	收藏品所在部室(華語)	收藏品所在部室(族語)
topic za	topic ind	type2	slang	zh tribe	ind tribe
topic zc	topic ind	type2	slang	zh tribe	ind tribe
topic zh	topic ind	type2	slang	zh tribe	ind tribe
topic zh	topic ind	type2	slang	zh tribe	ind tribe

Showing 1 to 4 of 4 entries

■ 族語新聞檢索

族語新聞檢索

k

顯示 10 結果

序號	視頻	標題	族語別	操作
1		海祭是噶瑪蘭最大的祭典	噶瑪蘭族	查看
2		擔心沒獵物 阿美'Masetik'避免尷尬	阿美族	查看

顯示第 1 至 2 項結果,共 2 項

上頁 1 下頁

族語新聞資訊

擔心沒獵物 阿美"Masetik"避免尷尬

Masetik



ano tayni i 'alo, ano tayni i
pariyaran ko salikaka itini a
misetik i,

如果去河邊，如果去海邊的同
胞在這裡抓捕撈

a. 播放影片一旁有逐字翻譯族語和中文的功能

文章 資訊

中文文章

如果去河邊，如果去海邊的同胞在這裡抓捕撈，他們常會與家裡的人說：「我去獵捕一下」，抓魚，抓螃蟹，男性不太會這樣直接說，來自靜浦部落(Ca'wi)的老人家不會那樣說，如果去河邊，如果去海邊，避免尷尬可能沒有帶東西回家則說「我去獵捕一下」，男性不太會說「我去捕魚」「我會去抓螃蟹」這樣子的話，在起身出發時，如果當時有人打噴嚏，過去來自長輩對禁忌的說法。

去捕撈，抓螃蟹去海邊捕魚，我去捕撈一下，去捕撈，抓螃蟹去海邊捕魚，我去捕撈一下，是... o sowe...去捕撈...是，我去抓魚...是，我去抓魚，我去捕撈一下，若要定義之... 如果沒有收獲，如果沒有那些魚，也沒有一隻的情形，我去捕撈一下這樣，像是螃蟹呀、蝦呀、魚呀，全部都被稱叫捕撈，捕魚，「我去捕魚一下」這樣，如果沒有魚，或者是蝦、其他或蟹，那就是那個，是我去獵捕一下，是「我去捕撈一下」這樣，去海邊，去河邊，Sa'owac部落那樣說，「我去捕魚一下」這樣，是長輩的事務，起程出發前，裝備是不可或缺的，起程出發開門之時，想打噴嚏之瞬間，表有不祥，還有其他的講法，如果我快達到時，在那時說「我去捕撈一下」，「我去捕撈一下」，我才達到這樣，不論是誰將打噴嚏，如果是小孩、或是長輩打噴嚏，都視為不祥、禁忌，既然是禁忌...宜謹慎小心。不管是什麼，若過於迷信其中，我們也不要心生恐懼，不管是

族語文章

ano tayni i 'alo, ano tayni i pariyaran ko salikaka itini a misetik i, ira kona sowal nangra, "masetik ho kako" sa itini a pakafana" to i loma'ay, caay itini a pafesoc a mifoting kako, a mikalang kako saan ca'ay kamatini ko sowal no fainayan. ca'ayay kona sowalanto no i Ca'wi a mato'asay a masadak
ano tayni 'alo, ano tayni i pariyaran, mangodo latek awa kona panokayan sa "masetik ho kako" saan, ca ko "mifoting kako," "o mamikalang kako" saan ca'ay kamatini ko sowal no fainayan. nika to kamo sa: dak tona lowad i, ano maesing, ano 'a'as ko tamdaw kiya i, ila ko sowal no mato'asay pakaynitini tona paysin.

masetik, mikalang talariyar mifoting. masetik, masetik ho kako. masetik, mikalang talariyar mifoting, masetik, masetik ho kako. hay, o sowe... masetik... hay mifoting ho kako sa... hay mifoting ho kako sa, masetik ho kako. ano pangangan han... ano awa ko kinaira, ano awa kora foting, awa pinang cecayay, masetik ho kako o sanay, ano o kalangto, ano o 'afarto, ano o fotingto. ma'emin macingangan tora setik, mifoting, mifoting ho kako saan, ano awa kiya ko foting iraira o 'afar ato maan ato kalang. oya hiya kira, hay masetik ho kako, hay masetik ho kako saan, talariyar, ano tala'alo, Sa'owac saan kora. masetik ho kako saan. o demak no mato'asay

b. 可比對閱讀文字

文章
資訊

新聞標題(華語)	擔心沒獵物 阿美"Masetik"避免尷尬
新聞標題(族語)	Masetik
新聞類別	語言文化
來源	原住民族電視臺
來源網站	https://www.youtube.com/watch?v=flfvjutuCl0
新聞日期	2018-09-30
最新修改時間	2018-12-13
轉譯者	無

c. 資訊呈現方式

■ 臺灣原住民族語言研究書目檢索

臺灣原住民族語言研究書目檢索

關鍵字
搜索

● 作者
● 年份
● 篇名
● 出版單位

您搜索 (篇名): 原住民, 共有 64 筆資料

顯示 結果
搜索:

序號	領域	文獻類別	篇名	操作
1	無	會議論文集 (有ISBN)	阿美族語言的巧與妙	查看

a. 檢索結果顯示在下方

b. 呈現欄位仍在協調中

c. 本年度將置入 2017 年語發中心彙整之臺灣原住民族語言研目錄彙編之資料。

■ 本站資訊檢索

Q

網站搜尋

目前仍在優化中，用以檢索本站頁面資訊。

■ 會員功能

會員登入

 請輸入郵箱

 請輸入密碼

[忘記密碼?](#)

含登入、註冊及找回密碼等功能

(3) 底部導覽列

 關於本站 <ul style="list-style-type: none">任務及簡介歷史沿革組織本站使用指南	 檢索 <ul style="list-style-type: none">典藏資料族語新聞臺灣原住民族語言研究書目	 提供資料 <ul style="list-style-type: none">資料收集內容資料提供方式資料管理計劃資料格式存取限制資料建置說明	 權利說明 <ul style="list-style-type: none">本站使用條件智慧財產權免責聲明授權同意書引用說明	 常見問題 <ul style="list-style-type: none">相關連結聯繫我們
--	--	---	--	---

■ 關於本站

任務及簡介

聯合國教科文組織 (UNESCO) 2009 年公布了「世界瀕危語言地圖」，其中臺灣有 5 種語言被列入極度瀕危，10 種語言被列入滅絕等級。因此除了進行語言振興之外，亦需要積極的保存語言。Crystal (2010:349) 提到保存語言的方式，最基本的是從語言記錄開始，從錄音 (聲音與影像)、分析、以文字記錄，這些都是為了教育。而這些調查的語料也應該要有一個平臺可以收錄。

我們查詢國內外族語保存相關的資料庫發現，國外早已有像是 AILLA、PARADISEC 這種網站，是提供收集多媒體語言資料、檔案的平臺，一般人都可以將所持有的語言保存資料提供給該網，就會數位化的被建置，甚至有分級訪問權限的制度，因此吸引許多研究人員或是族人願意提供資料。而這些資料也歡迎其他研究者協助分析、轉譯，讓資料庫收錄的越來越豐富。反觀國內，目前尚未有這種形式的資料庫，因

	<p>此這也是我們努力的方向。</p> <p>從歷史的角度來檢視，原住民族的語言開始有文字的記錄約莫是從荷據時期的「新港文書」開始，到日據時期之後有多位日本語言學者如小川尚義、淺井惠倫、土田滋、森口恒一、月田尚美、野島本泰為臺灣南島語言研究貢獻良多（李壬癸，2009）。而民國57年基督教聖經公會也以羅馬拼音將聖經翻譯成族語版，民國1992年則由教育部出版，委託由中央研究院李壬癸教授編訂臺灣南島語言的語音符號系統。臺灣原住民族語言雖因口傳方式保存不易，所幸尚有外來的傳教士、學者的記錄，使許多珍貴的語料尚得留存至今，但由於早年影音記錄及保存的技術尚不發達，可能造成部分資料因此損壞或遺失，十分可惜。</p> <p>爰此，如何應用現代技術，將已留存的族語文獻資料，特別是將早期的族語資料彙整，並以數位化的方式保存，是首要之務。將影音的部分加以轉譯成文字，以利後續研究或教學使用，便是未來我們需要努力的課題。故自2017年起我們便規劃建置「數位族語資料庫系統」，希望能以數位化方式保存原住民族語料，後續再擴大應用。</p>
歷史沿革	<p>依《原住民族語言發展法》第九條：「需建置原住民族語言資料庫，積極保存原住民族語料」，由原住民族委員會於2017年委託原住民族語言研究發展中心開始規劃，2018年開始進行第一期資料庫建置。</p> <ul style="list-style-type: none"> ◆ 2017年規劃 ◆ 2018年開始建置本站
組織	由原住民族委員會2017年始委託原住民族語言研究發展中心執行本計畫。
本站使用指南	指引閱覽大眾之指南，待網站開放前會修訂完成。

■ 檢索

典藏資料檢索	可檢索本站典藏之族語資料。
族語新聞檢索	本站每年會轉譯一定數量的族語新聞，可供檢索及查詢。
研究書目檢索	本站每年會增加研究書目，並提供檢索，瞭解目前研究的方向和概況。

■ 提供資料

資料收集內容	<p>為將珍貴的族語文獻資料彙集並進行數位典藏，以利資料之保存、研究及應用，且考量到授權問題，因此現階段語料主要是由研究人員以訪談方式收集語料，即以有音檔的口語語料優先，若有紙本資料再一併數位化收錄。除此之外，我們也安排族語新聞轉成文字的工作。</p>
資料提供方式	<p>目前本網站仍在先期建置階段，暫無開放公開上傳資料之計畫，若您有早期族語口說的影音資料願意提供，歡迎與我們聯繫。本網站先期建置內容類型，可分為：</p> <ul style="list-style-type: none"> ➢ 典藏資料：由各方言學者專家提供之長篇語料、錄音或錄影素材進行數位典藏。 ➢ 族語新聞：由原住民族文化事業基金會（原文會）所經營之原住民族電視臺播出之族語新聞（以影音頻道上如 Youtube 有儲存之新聞為主），再委託各族語專家學者把口語資料轉譯（繕打）為族語及翻譯為華語。 ➢ 研究書目：將有關臺灣原住民語言學相關之學術論文、期刊等研究著作建立目錄清單。
資料管理計畫	<p>目前我們管理收藏品的方式主要是將資料數位化後，將原版資料及一份備份光碟交還給貢獻者，並限制需同意遵守本站使用條款及尊重智慧財產權者才得以下載資料。</p>
資料格式	<p>為進行數位典藏，本站收錄的格式皆為數位化影音或文件格式。若取得的素材非數位格式，本站將會進行數位化保存，再列為本站族語典藏資源。您可以向本站連繫後再決定如何提供典藏的方式。</p> <ul style="list-style-type: none"> ➢ 典藏資料：包括紙本資料掃描檔案如 pdf, txt, jpg, png 等文件或圖檔格式；錄音帶轉檔格式如 mp3, mp4 等。
存取限制	<p>本站建置初期，內容不對外公開，僅開放部分建置內容(後設資料)。</p>

	<p>預計 2020 年後，視情況開放部分內容，以對外推廣，並將回饋之建議作未來資料庫建置大型計畫之參考。</p>
資料建置說明	<p>為考量人力及資源限制，以口語資料、瀕危語言及政府有版權之資料為優先建置標的，另資料之正確性亦十分重要，故目前擬先建置：</p> <ul style="list-style-type: none">➤ 族語新聞➤ 口訪語料➤ 政府委託計畫之語言調查成果 <p>以資料庫近期的規劃來說，預計在 2018 年完成資料庫前期的建置，主要功能會包含族語資料的數位典藏及檢索，另外也開使試辦族語新聞文字化工作。</p>

■ 權利說明

本站使用條件	<ul style="list-style-type: none">➤ 一般使用者（未註冊）：任何可上網的人都可以瀏覽本網站。可瀏覽和讀取本網站中資料，但無法存取任何檔案。使用者必須建立一個帳戶並登入才能瀏覽檔案。➤ 註冊的使用者：在本網站創建了帳戶並登入的使用者。除了能夠瀏覽典藏檔案和讀取後設資料之外，還可以查看或下載檔案。 <p>典藏資源的使用條件：</p> <p>本站要求使用者在建立帳號時同意典藏資源的使用條件。如果不同意，將無法建立帳號，也就無法存取本站任何典藏的檔案，包括指定為公開存取的文件。</p> <p>後設資料(metadata)：有關收集、檔案、語言、國家/地區、組織、儲存者(depositor)和數據的其他貢獻者的目錄或上下文信息。一般使用者和註冊使用者都可以存取後設資料。</p>
--------	---

	<p>檔案(files)：任何數位檔案（如：影像、聲音、圖片、文本等）。只有註冊使用者可以看到檔案，一般使用者無法查看或下載。</p> <p>公開存取(PublicAccess)：任何使用者都可以申請新的本站帳號，成為註冊的使用者後即可存取本站的檔案。公開存取和開放存取(OpenAccess)的不同之處在於開放存取意味著資源已被釋放到公眾領域中。由於大多數儲存於本站的檔案受版權保護，因此它們不是開放獲取。</p> <p>公開存取政策(PublicAccessPolicy)：本站鼓勵所有貢獻者提供族語資料（以口語為主），我們將協助您數位化，您可以當作將收藏品公開分享，但您還是保留您的著作權。</p> <p>限制存取(RestrictedAccess)：存取僅限於某些使用者帳戶，通常是收集或版權所有者或其代理人，以及本站管理人員。</p>
智慧財產權	<p>本站所有內容，都受到智慧創作條例之保護，我們歡迎您友善並在遵守法律原則下，尊重資料持有者的相關權利，本站之資源不得應用於商業使用，且任何引用皆需註明出處。</p> <p>若是您協助口語資料轉譯，或是有相關研究論文發表，還請回饋族人，將相關數位化檔案存放本站以供建置。</p>
免責聲明	<p>本資料庫是基於對原住民族語言資料數位化保存及分享的精神，善意且衷心期盼所有加入的成員皆能尊重檔案資源創作者的知識財產權，並且亦需尊重所載之族人和其語言、文化。如果有任何用戶違反了本資料庫檔案資源的使用條件，將可能面臨法律問題或社會的抨擊。本資料庫所保存的檔案，不得在未經過檔案資源所屬之部落或作者等同意逕自出版。</p> <p>關於檔案資源內容的免責聲明：</p> <ol style="list-style-type: none">1. 如果本中心必須對本網站內容中的錯誤或遺漏承擔任何責任或義務，則將無法提供

	<p>任何典藏資源或軟體。因此，語發中心對於因使用或無法使用本網站或使用典藏資源而導致的任何損失，不承擔任何責任。</p> <ol style="list-style-type: none">2. 同樣，如果本中心必須對任何典藏資源內容的任何部分的真實性或準確性負責，本中心網站就無法提供這些典藏資源，因此本網站不承擔任何此類責任。3. 任何典藏資源的內容或檔案資源提供者所表達的觀點、信仰或主張均不代表本網站的意見(或信仰)。4. 訪問本站的資源由該存儲者和創建者決定。可能會更改給定資源的訪問權限，恕不另行通知；本網站不對因此類變更而導致的任何損失或挫折承擔責任。5. 如果資源控制者將資源的存取限制從較低限制改為較高限制的級別，對於在較低限制期間可能已創建的任何存檔資源副本，本站不負任何責任。6. 一旦使用者存取資源並將其下載到他的電腦或其他存儲設備，使用者可能在未經許可的情況下複製和散佈資源，本站將無法掌控此類使用者行為。將典藏資源放在本站的人（提供者），必須對該行為的後果承擔全部責任。 <p>關於檔案資料來源的免責聲明：</p> <ol style="list-style-type: none">7. 如果本站對於適當的許可必須承擔責任，恐怕無法提供典藏資源。因此，提供材料的人必須負起全責來告知參與創建這些資源的所有人或其代表。8. 因典藏資源的存在或任何人對典藏資源的使用而造成任何可能的傷害或損失，本站不對下述事項承擔責任。如果本站必須接受此類責任，則可能無法提供典藏資源。9. 關於典藏資源的創建者和儲存者之間可能發生的衝突，本站不負責調解。對於介於創作者或儲存者、方言社區(的方言在這些資源中被使用或記錄)、或任何聲稱具
--	---

	<p>有代表性的方言組織之間的衝突，本站不負責調解。資源的儲存者必須在提交資源前討論並解決所有此類潛在衝突。</p> <p>10. 本站與創作者，資源提交者或方言組織之間的衝突，將交給本站諮詢委員會(或人員、團體之類的)進行調解。本站同意遵守董事會關於解決衝突的決定。</p> <p>關於網站和軟體的免責聲明：</p> <p>11. 在本網站上發表的意見並不代表語發中心的意見。您在本網站或任何存檔資源中找到的任何資訊的完整性、準確性或及時性，中心不對此負責。</p> <p>12. 本站不對連結到任何其他網站的內容負責，也不對任何提供連結到本站的網站的內容負責。</p> <p>13. 我們不保證任何應用軟體的可用性或安全性。對於相關軟體的使用風險必須由您自行承擔。</p>
授權同意書	當您提供實體資料請本網站作數位化保存時，請您填寫並提交紙本授權同意書，作

	在做出任何衍生作品之前請先請求授權（例如：轉錄、翻譯、納入教科書或分析、創作製作等）。
--	---

■ 常見問題

註冊	<p>您可以在不使用帳號的情況下瀏覽本網站各頁面，但如果沒有創建帳號並登入，則無法進行資料檢索、族語新聞檢索及研究書目檢索。</p> <p>創建帳戶時，您會收到來自 ilrdcLangAD@gmail.com 的電子郵件，要求您確認註冊。如果您沒有在收件箱中看到此電子郵件，請在垃圾收件夾中(或其他相類似之功能)尋找。按照電子郵件中的說明確認您的帳戶，您不用回復該電子郵件。</p>
忘記密碼	<p>如果您忘記密碼，您可以點選"登入"中的"忘記密碼"來申請新密碼。當您提交申請後將帶您到一個頁面，您將在其中輸入您的電子郵件地址(或用戶名)。提交請求後，您將收到來自 ilrdcLangAD@gmail.com 的有關修改密碼的電子郵件。</p> <p>如果您在收件箱中沒有看到此電子郵件，請在垃圾收件夾中(或其他相類似之功能)尋找。按照電子郵件中的說明登入並更改密碼。您不用回復該電子郵件。使用提供的鏈接登入後，您可以編輯您的帳戶資訊。參照密碼欄位旁的提示來加強您的密碼，然後在下一個框中再次輸入密碼以確認密碼。</p>
搜索與瀏覽	<p>本網站目前有三大搜索功能供您使用，您可以從本網站上方或下方連結點擊進入各搜索頁面，使用搜索欄搜尋族語、方言、地區、關鍵字等：</p> <p>➤ 典藏搜索：針對各族語的錄音或錄影，做採集、數位化(依來源格式而定)並典藏。</p>

	<ul style="list-style-type: none">➤ 族語新聞搜索：將各族的族語新聞整理，把口語資料轉譯為文字(中文及族語)，讓資料庫的族語口語資料更豐富。➤ 研究書目搜索：提供有關族語、語言學的研究資料蒐尋，包括各類論文、期刊等。
使用條款	<p>一般使用者(未註冊)：任何可上網的人都可以瀏覽本網站。可以瀏覽和讀取本網站中的後設資料，但他們無法存取任何檔案。要流覽檔案，使用者必須建立一個帳戶並登入。</p> <p>註冊的使用者：在本網站創建了帳戶並登入的使用者。除了能夠瀏覽典藏檔案和讀取後設資料之外，還可以查看或下載檔案。</p> <p>典藏資源的使用條件：本站要求使用者在建立帳號時同意典藏資源的使用條件。如果不同意，將無法建立帳號，也就無法存取本站任何典藏的檔案，包括指定為公開存取的文件。</p> <p>後設資料(metadata)：有關收集、檔案、語言、國家/地區、組織、儲存者(depositor)和數據的其他貢獻者的目錄或上下文信息。一般使用者和註冊使用者都可以存取後設資料。</p> <p>檔案(files)：任何數位檔案（如：影像、聲音、圖片、文本等）。只有註冊使用者可以看到檔案，一般使用者無法查看或下載。</p> <p>公開存取(PublicAccess)：任何使用者都可以申請新的本站帳號，成為註冊的使用者後即可存取本站的檔案。公開存取和開放存取(OpenAccess)的不同之處在於開放存取意味著資源已被釋放到公眾領域中。由於大多數儲存於本站的檔案受版權保護，因此它們不</p>

	<p>是開放獲取。</p> <p>公開存取政策(PublicAccessPolicy)：本站鼓勵所有貢獻者在本站免費提供其資料，並將本站典藏作為一種出版形式。對於內容和訪問，我們保留僅將我們認為合適的資源存檔的權利。</p> <p>限制存取(RestrictedAccess)：存取僅限於某些使用者帳戶，通常是收集或版權所有者或其代理人，以及本站管理人員。本站有數種不同類型的限制存取規定。</p>
下載	<p>您必須登入才能存取本網站典藏的任何檔案。如果您已登錄但仍無法存取檔案，則可能需要調整瀏覽器設置以接受 cookie。網站上的大部分資源都可以由公開無受限地存取，但有些資源受到提交者的限制。這些限制可以保護受訪者免受傷害或尷尬。有時您會被要求直接聯繫提交者以獲得存取該檔案的授權。這使提交者能夠知道誰在使用他們的資源。</p> <p>除非您的網路速度非常快，否則某些檔案可能無法下載。您可以在開始下載之前查看媒體詳細資訊以了解檔案大小。</p> <p>一些未壓縮的歸檔檔案（wav，mpg 和 tif）即使具有後設數據記錄也無法及時獲取。這些檔案在資源頁面的檔案類型下沒有列出任何內容，並且該檔案的記錄頁面上沒有下載按鈕。很抱歉給您帶來不便。如果情況允許，請下載壓縮過的等效檔案（mp3，mp4，jpg 或 pdf）。</p>
查看與播放檔案	<p>本站的許多檔案-聲音、影像、圖像和文件檔案，可以在瀏覽器中直接查看或串流。您還可以使用可以免費的程式下載檔案並在您自己的設備上播放它們。</p> <p>➤ 聲音（wav, mp3）和影像（mpg, mov, avi, mp4, m4v 等）檔案：iTunes, Quicktime,</p>

	<p>Windows Media Player, Audacity, VLC 媒體播放器等。</p> <ul style="list-style-type: none">➤ 圖像 (jpg, png, gif, tif) : Windows Photo Viewer, iPhoto, IrfanView 等。➤ PDF 檔案 : Adobe Acrobat Reader。➤ 純文字 (txt) 檔案 : Wordpad, TextWrangler, Notepad, Word, Open Office 等。注意：如果缺少所有換行(文件不如預期情況下顯示斷行)，您可能正在嘗試使用記事本。請嘗試用不同的程式開啟，可能會解決您的問題。記事本有時無法正確轉換在其他平臺上創建的文字檔案。➤ Transcriber (trs) 字幕檔案 : Transcriber 程序生成的轉錄，翻譯和註釋檔案，請參考純文字檔案。
使用與引用	<p>引用任何資源前，您都必須遵守本站的引用指南（包括但不限於教育材料、演講稿、出版品、展覽）。本網站的資源可以用於以下用途：</p> <ul style="list-style-type: none">➤ 為所有年齡和課堂開發教材。➤ 用於研究。➤ 藝術創作。➤ 其他。 <p>我們鼓勵您探索檔案並充分利用這些資源，但我們要求您始終尊重：</p> <ul style="list-style-type: none">➤ 創作這些材料的個人。➤ 在這些材料中說話(發言)的人。➤ 於本網站產出和儲存資源的學者們。 <p>在做出任何衍生作品之前請先請求授權（例如：轉錄、翻譯、納入教科書或分析、創作製作等）。</p>

■ 相關連結

國內相關網站：

- 臺大臺灣南島語多媒體語料庫
http://corpus.linguistics.ntu.edu.tw/index_zh.php
- 臺灣南島語數位典藏（目前已停用）
<http://languagearchives.sinica.edu.tw/cht/index.php.html>
- 蘭嶼達悟語口語資料典藏
http://yamiproject.cs.pu.edu.tw/yami/yami_ch/ch_index_flash.htm
- 達悟語線上學習網：
<http://yamiproject.cs.pu.edu.tw/elearn/?&lang=zhtw>
- 蘭嶼媒體與文化數位典藏：<http://lanyu.nctu.edu.tw/database/>
- 中央研究院語言學所「典藏計畫」之臺灣南島語數位典藏（目前已停用）：
<http://languagearchives.sinica.edu.tw/cht/index.php.html>。
- 臺灣原住民族數位典藏資料庫：
<http://dore.tacp.gov.tw/tacp/pingpu/index.html>
- 臺灣聖經公會網站：<http://cb.fhl.net/>
- 臺灣語言學會（語典及語料庫）：<http://linguist.tw/zh-tw/links/%E7%9B%B8%E9%97%9C%E7%B6%B2%E8%B7%AF%E8%B3%87%E6%BA%90%E9%80%A3%E7%B5%90/dictionaries%20and%20corpus>。
- 花蓮縣國小本土語文領域輔導團（太魯閣字典）：
<http://teacher.hlc.edu.tw/imax9.asp?id=81&sid=259>
- 阿美族、雅美族口語傳說採錄翻譯資料數位典藏計畫：
<http://aya.ioe.sinica.edu.tw/>
- 阿美語萌典：<https://amis.moedict.tw/>
- 阿美語線上讀經：http://www.amis100.org.tw/cudad_0_0.htm
- 政大原民中心電子書城：<http://ebook.alcd.tw/>
- 數位典藏國家型計畫—臺灣社會人文電子影音數位博物館計畫：<http://proj1.sinica.edu.tw/~video/search/index1.html>
- 教育部朗聲四起（原住民族語讀文章）：<https://alcd-con.tw/alr/lobby.php>
- 創用 CC 授權條款：<http://creativecommons.tw/license>

國外相關網站：

- AILLA: The Archive of the Indigenous Languages of Latin America，拉丁美洲原住民族語言檔案館 <https://ailla.utexas.org/>
- Daly Languages AUSTRALIA，澳洲 Daly 語語料庫 <http://www.dalylanguages.org/>
- Glottolog，世界語言參考文獻網 <http://glottolog.org/>
- PARADISEC: Pacific and Regional Archive for Digital Sources in Endangered Cultures，太平洋地區檔案數位資源瀕危文化庫 <http://www.paradisec.org.au/>
- ASEDA (The Aboriginal Studies Electronic Data Archive)，澳洲原住民研究電子數據檔案庫：<http://aseda.aiatsis.gov.au>
- Corpus of Contemporary American English，美國當代英語料庫：<https://corpus.byu.edu/>
- Kevin Scannell on Corpus-based language technology for indigenous and minority languages：
<https://www.youtube.com/watch?v=W05YmEl093o>
- LDC (Linguistic data Consortium, University of Pennsylvania)，美國賓州大學語言資料庫：
<https://www ldc.upenn.edu/>
- OLAC: Open Language Archives Community，開放語言檔案社群網：<http://www.language-archives.org/>
- UNESCO Interactive Atlas of the World's Languages in Danger
聯合國教科文組織世界瀕危語言地圖：
<http://www.unesco.org/languages-atlas/index.php>

■ 連繫我們

Figure 4 shows two contact forms side-by-side. The left form is titled '聯繫我們' (Contact Us) and includes fields for '姓名' (Name), '聯絡電話' (Contact Number), '郵箱' (Email), '主題' (Subject), and '問題描述(500字內)' (Problem Description, 500 characters). The right form is also titled '聯繫我們' and includes fields for '主題' (Subject) and '問題描述(500字內)' (Problem Description, 500 characters). Both forms have a '提交' (Submit) button. A box above the right form contains the text '分為會員及非會員介面' (Divided into member and non-member interfaces).

(二) 資料庫系統功能（後臺）

基本後臺系統功能如下圖：

Figure 5 is a screenshot of an administrator backend interface. It features a central '功能管理' (Function Management) section with a list of functions: '個人檔案' (Personal Profile), '密碼更改' (Change Password), '會員管理' (Member Management), '檔案管理' (File Management), '逐字稿管理' (Transcript Management), '族語新聞管理' (Indigenous Language News Management), and '貢獻者管理' (Contributor Management). To the right, there is a list of other system functions: '典藏資料管理' (Collection Management), '消息公告管理' (Message Management), '逐字稿編輯' (Transcript Editing), '族語新聞編輯' (Indigenous Language News Editing), '台灣原住民族語言研究書目系統' (Taiwan Indigenous Language Research Bibliography System), and '問題回報管理' (Problem Reporting Management).

而目前資料庫系統權限及功能如下：

- 一般使用者：個人檔案、密碼更改。
- 受指派工作者：典藏資料、族語新聞（逐字稿及族語新聞編輯）、研究書目、消息公告、問題回報。
- 管理者：會員管理、檔案管理、逐字稿管理、族語新聞管理、貢獻者管理、典藏資料庫理、消息公告管理、臺灣原住民族語言研究書目系統、問題回報管理。

※以下依序介紹資料庫系統功能：

1. 帳號系統

(1) 註冊：



帳號申請

Email *
example@gmail.com

密碼*
密碼確認*

姓名*
王小明

族名

族別* 方言別* 部落別
請選擇 請選擇 請選擇

出生日期
mm/dd/yyyy

性別* 女 男

身分證統一編號*

- 填入完整資料後便能註冊成功
- 註冊成功後會發送認證連結至該信箱進行認證
- 認證成功後資料庫將轉變帳號狀態，便能正常登入
- 認證連結有效期限為一天

(2) 登入：從網站的左上角的按鈕登入



會員登入

請輸入Email

請輸入密碼

登入

註冊

忘記密碼?

- 登入有效期限為七天

(3) 忘記密碼：

密碼重設

Email *
example@email.com

提交

密碼重設

密碼*
密碼確認*
提交

- f. 在網址中輸入電子信箱
- g. 若資料庫有該帳號，則發送修改密碼連結至該信箱
- h. 密碼修改成功後，請使用新密碼登入
- i. 修改密碼連結有效期限為一天

2. 新聞逐字稿系統

(1) 族語新聞管理*：

a. 可查看所有的項目

族語新聞

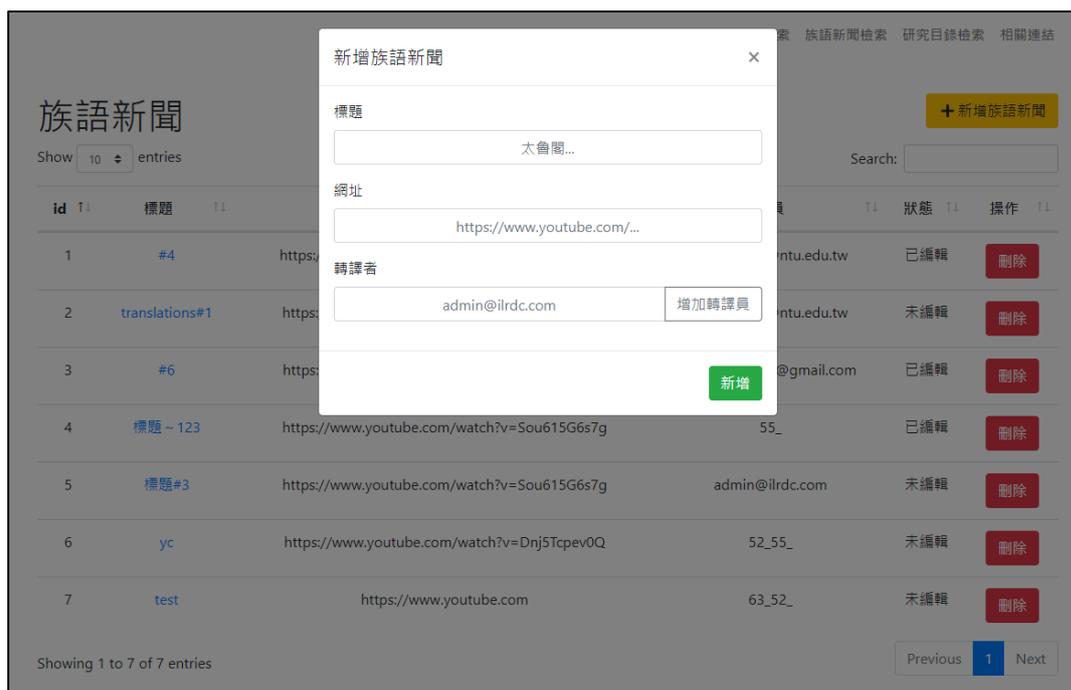
Show 10 entries Search:

id	標題	網址	翻譯員	狀態	操作
1	#4	https://www.youtube.com/watch?v=HhuNyfw-a6M	b03502151@ntu.edu.tw	已編輯	刪除
2	translations#1	https://www.youtube.com/watch?v=Sou615G6s7g	b03502151@ntu.edu.tw	未編輯	刪除
3	#6	https://www.youtube.com/watch?v=gc77DX9iGC8	yincheng0819@gmail.com	已編輯	刪除
4	標題 ~ 123	https://www.youtube.com/watch?v=Sou615G6s7g	55_	已編輯	刪除
5	標題#3	https://www.youtube.com/watch?v=Sou615G6s7g	admin@ilrdc.com	未編輯	刪除
6	yc	https://www.youtube.com/watch?v=Dnj5Tcpev0Q	52_55_	未編輯	刪除
7	test	https://www.youtube.com	63_52_	未編輯	刪除

Showing 1 to 7 of 7 entries

Previous 1 Next

b. 可新增項目



c. 可刪除項目

(2) 逐字稿系統*：



- d. 可進行逐字稿轉譯作業
- e. 可擷取時間
- f. 選定影片範圍進行播放
- g. 有快捷鍵功能

(3) 新聞編輯系統*：

族語新聞編輯

顯示 結果 搜索:

序號	類別	標題	族別	方言	逐字稿	文稿狀態	最新編輯日期
4	語言文化	海祭是噶瑪蘭最大的祭典	噶瑪蘭族	無	查看	已完成	2018-12-10 09:53:47
5	語言文化	擔心沒獵物 阿美"Masetik"避免尷尬	阿美族	海岸阿美語	查看	已完成	2018-12-13 16:31:25
1			排灣族	無	查看	未完成	2018-12-13 16:31:04
2			太魯閣族	無	查看	未完成	2018-11-23 13:51:40
3			阿美族	海岸阿美語	查看	未完成	2018-11-30 14:55:45
6		獵熊需抵命 泰雅獵人視為禁忌 2018-07-05 A...	泰雅族	賽考利克泰雅語	查看	未完成	2018-12-15 14:44:21
7			布農族	卓群布農語	查看	未完成	2018-12-13 16:28:35

顯示第 1 至 7 項結果,共 7 項 上頁 下頁

族語新聞編輯

[播放影片](#) [查看逐字稿](#)

新聞日期:

新聞標題(華語):

新聞標題(族語):

新聞類別:

關鍵字:

來源:

來源網站:

族語文章:

中文文章:

- a. 可進行編輯標題、族別、方言等資訊
- b. 將逐字稿內容編輯成文章
- c. 完成後前臺即可見

*為需登入帳號

3. 典藏資料

(1) 典藏資料管理：

典藏資料管理系統

+ 新增典藏資料

全部勾選 共選擇了0項, 執行操作: 刪除 匯出

顯示 10 結果

搜索:

序號	類別	語別	方言別	主題	年代	內容簡述	狀態	操作	操作	操作
<input type="checkbox"/> 1	訪談	太魯閣族	太魯閣語	七、漫談民初太魯閣事件	1970-01-01-1970-01-01	漫談民初的太魯閣事件	已完成	編輯	刪除	匯出
<input type="checkbox"/> 2	訪談	太魯閣族	太魯閣語	五、灰燼	1970-01-01-1970-01-01	灰燼	已完成	編輯	刪除	匯出
<input type="checkbox"/> 3	訪談	太魯閣族	太魯閣語	六、手工藝	1970-01-01-1970-01-01	手工藝	已完成	編輯	刪除	匯出
<input type="checkbox"/> 4	訪談	太魯閣族	太魯閣語	一、獻祭品	1970-01-01-1970-01-01	獻祭品	已完成	編輯	刪除	匯出
<input type="checkbox"/> 5	訪談	太魯閣族	太魯閣語	二、彩虹	1970-01-01-1970-01-01	彩虹	已完成	編輯	刪除	匯出
<input type="checkbox"/> 6	訪談	太魯閣族	太魯閣語	三、狩獵	1970-01-01-1970-01-01	去野外 (狩獵)	已完成	編輯	刪除	匯出

<http://fsecure.cc.ntu.edu.tw:3000/collection/search>

- 可進行資料檢索及排序
- 可進行資料編輯、刪除、匯出功能

(2) 典藏資料上傳：

新增典藏資料

首頁 最新消息及公告 典藏資料檢索 族語新聞檢索 研究目錄檢索 相關連結

收藏主題(華語)

收藏主題(族語)

族語別

方言別

收藏器網站

收藏者名稱(中文)

收藏者名稱(族語)

其他收藏者名稱(列出收藏者的名稱)

收藏品所在國家

收藏品所在縣市

收藏品所在鄉鎮

收藏品所在村里

收藏品所在部落(華語)

收藏品所在部落(族語)

收藏品年代

收藏品內容描述(華語)

收藏品內容描述(族語)

- 輸入典藏詳細資料
- 可上傳文檔，預設為不可下載
- 輸入資料好後，送出將資料存入

(3) 檔案管理：

檔案管理							
全部勾選 共選擇了0項，執行操作： 刪除							
顯示 <input type="text" value="10"/> 結果						搜索: <input type="text"/>	
序號	上傳者	檔案名字	上傳時間	檔案類別	檔案大小	下載	操作
<input type="checkbox"/>	1	ILRDC	arrow-down.svg	2018-09-20 17:10:12	image/svg+xml	591 Bytes	下載 刪除
<input type="checkbox"/>	2	鄧永泰	calendar.svg	2018-09-20 18:04:56	image/svg+xml	2 KB	下載 刪除
<input type="checkbox"/>	3	鄧永泰	beaker.svg	2018-09-20 18:07:04	image/svg+xml	1 KB	下載 刪除
<input type="checkbox"/>	4	鄧永泰	bold.svg	2018-09-20 18:07:53	image/svg+xml	940 Bytes	下載 刪除

a. 管理已上傳之檔案，可搜索、下載及刪除

檔案編輯

	tlrdc文本7漫談民初太魯閣事件.pdf (5 MB)
檔案提供者	✕ 黃堂維 +
檔案語言別	太魯閣族語 <input type="text"/>
檔案族言別	太魯閣語 <input type="text"/>
檔案其他語言	
檔案內容類型	
連絡人姓名	
連絡人信箱	
檔案上傳時間	2018-11-01 上午11:29
檔案修改時間	2018-11-16 上午10:30

開放下載檔案

[提交](#)

b. 編輯檔案路徑為典藏資料 > 點繫欲修改檔案 > 檔案編輯

c. 在此編輯檔案是否開放下載

4. 研究書目管理

臺灣原住民族語言研究書目管理

+ 新增

關鍵字 搜索 列出所有

● 作者 ● 年份 ● 篇名 ● 出版單位

顯示 10 結果

搜索:

序號	領域	文獻類別	篇名	狀態	操作	操作
1	語言學領域 (一般語言學、應用語言學)	學位論文	無	已完成	編輯	刪除
2	語言學領域 (一般語言學、應用語言學)	學位論文	無	已完成	編輯	刪除
3	語言學領域 (一般語言學、應用語言學)	學位論文	無	已完成	編輯	刪除
4	語言政策、法規領域 (法律政策：政策、語言權、語言規劃、師資培育政策)	學位論文	無	已完成	編輯	刪除

臺灣原住民族語言研究書目新增

領域	*文獻類別
<input type="text" value="請選擇"/>	<input type="text" value="請選擇"/>
研究之族語別	研究之方言別
<input type="text" value="請選擇"/>	<input type="text" value="請選擇"/>
研究之語文別	作者族別
<input type="text" value="請選擇"/>	<input type="text" value="請選擇"/>
作者	書集編輯(著)
<input type="text"/>	<input type="text"/>
出版年	篇名
<input type="text" value="EX: 2018"/>	<input type="text"/>
起迄頁數	出版地
<input type="text"/>	<input type="text"/>
出版單位/者	中文關鍵詞
<input type="text"/>	<input type="text"/>
英文關鍵詞	中文摘要
<input type="text"/>	<input type="text"/>

- a. 可新增、編輯、刪除及檢索書目
- b. 欄位待調整

(三) 已建置之資料

1. 典藏資料：

已上傳 2 篇敘事訪談、10 筆長篇語料

典藏資料管理系統

[+ 新增典藏資料](#)

全部勾選 共選擇了0項。 執行操作: [刪除](#) [匯出](#)

顯示 [25] 結果 搜索:

序號	類別	語別	方言別	主題	年代	內容簡述	狀態	操作	操作	操作	
<input type="checkbox"/>	1	訪談	太魯閣族	太魯閣語	七、濃談民初太魯閣事件	1970-01-01-1970-01-01	濃談民初的太魯閣事件	已完成	編輯	刪除	匯出
<input type="checkbox"/>	2	訪談	太魯閣族	太魯閣語	五、灰煙	1970-01-01-1970-01-01	灰煙	已完成	編輯	刪除	匯出
<input type="checkbox"/>	3	訪談	太魯閣族	太魯閣語	六、手工藝	1970-01-01-1970-01-01	手工藝	已完成	編輯	刪除	匯出
<input type="checkbox"/>	4	訪談	太魯閣族	太魯閣語	一、獻祭品	1970-01-01-1970-01-01	獻祭品	已完成	編輯	刪除	匯出
<input type="checkbox"/>	5	訪談	太魯閣族	太魯閣語	二、彩虹	1970-01-01-1970-01-01	彩虹	已完成	編輯	刪除	匯出
<input type="checkbox"/>	6	訪談	太魯閣族	太魯閣語	三、狩獵	1970-01-01-1970-01-01	去野外 (狩獵)	已完成	編輯	刪除	匯出
<input type="checkbox"/>	7	訪談	太魯閣族	太魯閣語	四、變成啞吧的狗	1970-01-01-1970-01-01	變成啞吧的狗	已完成	編輯	刪除	匯出
<input type="checkbox"/>	8	訪談	太魯閣族	太魯閣語	八、德國孩子的成長歷程	1970-01-01-1970-01-01	德國孩子的成長歷程	已完成	編輯	刪除	匯出
<input type="checkbox"/>	9	訪談	太魯閣族	太魯閣語	九、狗變成啞巴	1970-01-01-1970-01-01	狗變成啞巴	已完成	編輯	刪除	匯出
<input type="checkbox"/>	10	訪談	太魯閣族	太魯閣語	十、思跟阿威和思跟希立	1970-01-01-1970-01-01	思跟阿威和思跟希立	已完成	編輯	刪除	匯出
<input type="checkbox"/>	11	其他	太魯閣族	太魯閣語	梨子的故事	1970-01-01-1970-01-01	梨子的故事,看影片用全...	已完成	編輯	刪除	匯出
<input type="checkbox"/>	12	其他	太魯閣族	太魯閣語	青蛙的故事	1970-01-01-1970-01-01	青蛙的故事,觀看沒有文...	已完成	編輯	刪除	匯出

顯示第 1 至 12 項結果,共 12 項 上頁 下頁

2. 族語新聞及逐字稿 5 則：

族語新聞編輯

顯示 [10] 結果 搜索:

序號	類別	標題	族別	方言	逐字稿	文稿狀態	最新編輯日期
1	語言文化	講求“文如其分”排灣族文身手法解祕!	排灣族	無	查看	已完成	2018-12-17 04:36:30
3	語言文化	阿美婚喪喜慶結束 Pakelang完工管謝	阿美族	海岸阿美語	查看	已完成	2018-12-17 04:40:18
4	語言文化	海祭是噶瑪蘭最大的祭典	噶瑪蘭族	無	查看	已完成	2018-12-10 09:53:47
5	語言文化	擔心沒獵物 阿美“Masetik”避免尷尬	阿美族	海岸阿美語	查看	已完成	2018-12-13 16:31:25
7	語言文化	布農族主要有6個氏族 以玉山為中心建立族群	布農族	卓群布農語	查看	已完成	2018-12-17 04:38:43

四、計畫執行成效

綜合以上的執行內容，我們從族語語料清單檢視、辦理研商會議、數位化語料、辦理轉譯工作等，建置並完成第一期數位族語資料庫，本年度的工作主要成效如下：

- (一) 數位化 39 卷族語語料錄音帶（其中 1 卷原帶損壞待修復）。
- (二) 辦理 9 場研商會議，包含資料庫建置工作、族語新聞轉譯工作等。
- (三) 建置第一期原住民族語言研究發展中心數位族語資料庫，內含基本頁面、公告系統、會員管理系統、檔案管理系統、逐字稿系統、族語新聞編輯系統、臺灣原住民族語言研究書目系統、問題回報及管理系統等。而資料庫首頁亦正在進行標題族語化工作，匯整。
- (四) 提出「數位族語資料庫第一期建置內容及規劃修正報告」，包含建置內容、階段性建置任務及目標修正。

肆、數位族語資料庫修正規劃

我們以 2017 年「數位族語資料庫系統的第一階段的規劃」為基礎，並參考資料庫研商修正會議意見，提出 2019 年資料庫的維護與建置方向，並修正階段性建置任務及目標。

一、資料庫建置規劃修正

(一) 增加各族語言介面切換功能

1. 可提升族語的能見度，並鼓勵族語的使用，亦符合本站設立之目的及精神。
2. 目前已先翻譯 15 族首頁標題項目（如附錄）。
3. 未完成的族別繼續完成翻譯工作。

(二) 典藏資料功能調整

1. 資料集合可以以提供者（目前稱貢獻者）為單位，以澳洲 paradisec 站的方式做修正。
2. 建置資料項目內容需再調整。
3. 檢索功能可加入類別式或進階式檢索。
4. 上傳文件應加入版權宣告。
5. 檔名格式可修正為自動編碼。
6. 未來可結合地理資訊系統

(三) 族語新聞功能

1. 未來可以參考 VoiceTube 英文學習網站功能，將語音、文字結合族語教學詞庫，例如學習詞表、教材。
2. 逐字稿功能未來也可考慮開發的更完善，可參考 youtube 功能。

(四) 次網頁檢視

1. 版型需再做調整。

2. 需增加修改頁面功能。

(五) 會員功能

1. 需將個人檔案之個人資訊部分，如身分證字號、生日以星號替代，以保障資料安全。

(六) 語料建置

1. 優先處理已收集之資料，將之數位化並進行修復工作。
2. 品質完好之音檔，優先進行內容整理，並盡快請教耆老，補充檔案資訊，典藏於系統。

(七) 維運備份

1. 需將資料庫原始碼製成光碟，以利還原。
2. 備份除了伺服器外，也應定期備份於中心硬碟。
3. 災變修復技術應回歸由中心技術人員負責，以避免外包團隊無法繼續協助無人可修復之情形。

(八) 資料庫內容架構

目前規劃修正的資料庫的架構如下圖：

資料庫近期規劃			資料庫未來發展	
主要功能 <ul style="list-style-type: none"> ◆典藏族語資料並提供檢索功能 	資料庫系統建置 <ul style="list-style-type: none"> ◆資料庫系統委外建置 ◆系統修正 	資料典藏建置 <ul style="list-style-type: none"> ◆將經整理及確認後資料建置於系統 	未來發展功能 <ul style="list-style-type: none"> ◆應用於教學、發展語料庫 	口語語料庫 <ul style="list-style-type: none"> ◆開發轉譯輸入系統 ◆口語語料轉譯 ◆語料標記與分析
資料形式 <ul style="list-style-type: none"> ◆影音 ◆書面資料 	資料保護機制 <ul style="list-style-type: none"> ◆重視智慧財產權及研究倫理議題 ◆取得授權及授權使用機制 	資料庫部分開放 <ul style="list-style-type: none"> ◆開放檢索功能，下載檔案需註冊取得使用權限 ◆網站有勘誤回報之聯繫方式 	未來可擴充建置之資料 有族夠人力及資源，並有其他實際使用之需求，可擴充建置之資料如下： <ul style="list-style-type: none"> ◆出版品 ◆政府相關計畫可供建置之語言調查成果語料(如與推人員、瀕危計畫、族與保母計畫等) ◆私人收藏 ◆族語新聞(持續增加) ◆族語研習課程作業 ◆族語認證測驗答卷及口語 ◆政府族語公文書 	學習者語料庫 <ul style="list-style-type: none"> ◆開發轉譯輸入系統 ◆族語測驗語料轉譯 ◆語料分析與標記
優先建置資料 考量人力及資源限制，以口語資料、瀕危語言及政府有版權資資料為優先建置標的，另資料之正確性亦十分重要，故目前建置： <ul style="list-style-type: none"> ◆族語新聞 ◆口訪語料 ◆原民會委託計畫可供建置之語言調查成果語料 	資料數位化 <ul style="list-style-type: none"> ◆書面資料數位化建檔，依資料可先掃描若資料狀態良好，可使用OCR光學字元識別 ◆口語資料，影音檔先行數位化，分批整理，供未來口語資料庫使用 	調整與修正 <ul style="list-style-type: none"> ◆召開研商會議及依據使用者回饋意見，調整系統 		其他語料庫 <ul style="list-style-type: none"> ◆依需求及資料庫資料進一步發展的語料庫
				研究應用 <ul style="list-style-type: none"> ◆研究書目網站 ◆原住民族語言研究語料彙集
				教學應用 <ul style="list-style-type: none"> ◆學習者語言使用 ◆教材編輯運用 ◆常用詞彙與搭配詞 ◆詞彙難易度分級，使語言能力測驗、教材有較明確的分級標準

圖 6 原住民族語言研究發展中心族語資料庫建置規劃架構

以資料庫近期的規劃來說，預計在 2019 年完成資料庫中期建置，主要功能是完善資料庫典藏及檢索功能，而典藏資料主要以未有版權問題，並以口語資料及瀕危語言為優先。資料的形式也會有口語及書面資料，進行數位化的方式在書面資料的部分，會先掃描為 PDF 檔及嘗試使用 OCR 光學字元識別，但仍有可能有缺字情形，尚需人工校對，而口語資料則依人力及資源分批進行轉譯工作，將之文字化及數位化，並請族人協助來視，完成前置工作則進行資料庫建檔，以利資料之典藏及應用。

(九) 智慧財產權與授權

我們參考 AILLA¹網站，使用者於申請創建²帳號時就應同意網站使用的條款，內容主要是不能將資源用於商業目的、尊重網

¹ AILLA (The Archive of the Indigenous Languages of Latin America)，拉丁美洲原住民族語言檔案館，網址：<https://ailla.utexas.org/>

² AILLA 註冊頁面及條款 <https://ailla.utexas.org/user/register>

站的知識權和版權，而未經合法授權不得修改資料庫中的任何資源、也不得以資料庫中的資料去創作其他作品，善意的使用系統及尊重分級權限的限制，若下載的元數據資源中創作者和參與者的名字需匿名，不管口頭或是書面形式都要尊重他們的匿名，最後是會根據引用指南³引用在本站使用的資源。另外也提到，本站有很多未經轉錄、未經翻譯及未經分析的錄音，很歡迎用戶協助將口語資料翻譯及分析，提供貢獻，條款原文如下圖。

I have read and accept the terms and conditions listed on this page. *
Terms and Conditions for Use
Resources deposited in the AILLA database are protected by the AILLA License Agreement or by more specific agreements signed by the depositors and AILLA's directors. While AILLA can not guarantee that no violations of these agreements will occur, if we learn of a violation, AILLA will notify the AILLA community of the violation.

"I agree to the following conditions on the use of AILLA resources:

1. I will not use any resource in the AILLA database for commercial purposes.
2. I will respect the intellectual property rights and copyrights that pertain to AILLA's holdings.
3. I will not alter or modify any resource in the AILLA database without a legal authorization to do so, for example, explicit permission from the rights holder or Fair Use of Copyrighted Materials (US Code Title 17, Sec. 107).
4. I will not create works derived from any resource in the AILLA database, without a legal authorization to do so, for example, explicit permission from the rights holder or Fair Use of Copyrighted Materials (US Code Title 17, Sec. 107).
5. I will act in good faith in interactions with AILLA's User Registration and Graded Access Systems. This means that I will not attempt to disguise my electronic identity or impersonate an AILLA depositor or resource creator, and I will abide by the decisions made by resource controllers concerning access to AILLA resources.
6. If the metadata for a resource states that names of creators and participants must be kept anonymous, I will respect their anonymity in any spoken or written representation of that resource that I produce.
7. I will cite any AILLA resource that I make use of in any published work according to the AILLA Citation Guidelines."

If you produce transcriptions or translations or other derivative materials from AILLA resources that our user community would find interesting or useful, we would appreciate it if you would deposit those materials with us as well. AILLA has many recordings that are untranscribed, untranslated, and unanalyzed, and we hope that members of our user community will contribute to AILLA by helping to make these oral works more accessible with transcriptions, translation and analyses.

Thank you for your interest in our collection.

圖 7：AILLA 網站使用條款和條件，需同意才能註冊

目前網站尚未大規模開放民眾自行上傳，而是透過本站工作人員優先處理所收集之語料，而針對要建置的資料，我們基於研究倫理及智慧財產權，皆會簽訂調查知情同意及相關授權書。

二、階段性建置任務及目標

按原住民族語言振興第 2 期六年計畫的規劃，語發中心預期任務是到 2019 年，而未來財團法人原住民族語言研究發展基金會（下稱原語發基金會）成立後，相關計畫應由該會繼續執行。因此我們先 2019 中期的族語資料庫建置內容，如下圖：

³ AILLA 引用指南說明 <https://islandora-ailla.lib.utexas.edu/site/rights/citation>



圖 8：數位族語資料庫階段性建置任務及目標

如上圖所呈現，2019 年主要是繼續資料庫系統功能的修正，並繼續增加族語語料，待原語發基金會成立後，將系統移轉。而 2020 年後，則應整合相關資源，擴大資料庫的規模及發展應用功能。

伍、結論與建議

提供語言資料的保存，就是「建置數位族語資料庫」最核心的目的。本計畫廣續 2017 年的規劃，今年已建置第一期的數位族語資料庫網站及系統，並將數位化的口語語料、文本及族語新聞、研究書目等逐一進行測試及建置於系統中。

由於族語文字化仍可能面臨到書寫系統尚未修訂完成，以及書寫規範尚未有共識的問題，故在族語新聞或口語語料轉寫時，我們的原則有二，首先書寫符號使用 2005 年頒布版本，其次至少同一人使用的書寫規則前後一致，以利後續修訂版本公告時，較容易以批次修改的方式修改。

今年資料庫中，完成編校的資料預計有族語新聞 5 則、典藏資料 12 篇（含長篇語料經掃瞄光學辨視檔 10 筆、故事敘事 2 篇），研究書目的部分另於本中心研究書目報告中呈現。

為了資料庫建置修正及維護工作更快銜接，我們試針對 2019 年工作提出規劃：

一、2019 年工作規劃

（一）執行內容

1. 依語發中心 2018 年「數位族語資料庫第一期建置內容及規劃修正報告」為基礎，修正數位族語資料庫系統部分功能。
2. 將獲得授權之資料繼續數位化，並進一步訪談及整理，將完成資料建置於系統。
3. 辦理資料庫研商會議。
4. 提交「108 年數位族語資料庫建置報告」1 種，內容應含：
 - （1）建置內容說明。
 - （2）階段性任務分析。

(二) 工作流程

2019 年工作流程安排如下圖：

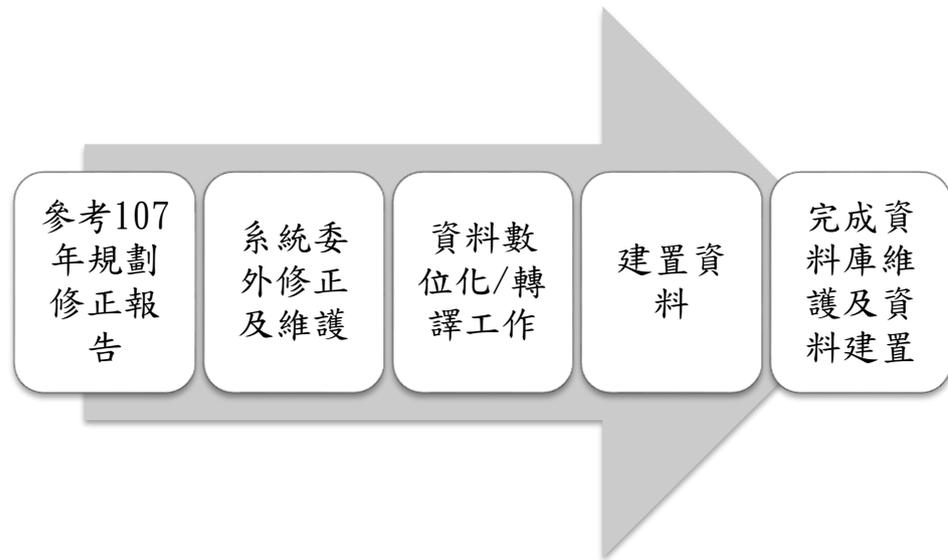


圖 9：2019 年語發中心「數位族語資料庫維護與資料建置」工作規劃流程

1. 參考 107 年規劃修正報告
 - (1) 依修正報告內容提出系統修改細節需求書。
 - (2) 辦理研商諮詢及工作會議。
2. 委外進行系統修正及維護工作
 - (1) 系統修正需求書。
 - (2) 委外進行資料庫介面及系統修正工作。
3. 資料數位化及進行轉譯工作
 - (1) 資料進行數位化及修復。
 - (2) 請專業人才協助族語新聞進行轉譯及編輯工作。
4. 完成資料庫維護及資料建置
 - (1) 確保計畫期間資料庫之運作。
 - (2) 本年度建置 12 則族語新聞及 5 卷語料。
 - (3) 提出本年度建置內容報告。

(三) 工作時程規劃

2019 年工作時程規劃預計如下：

工作項目\月份	1	2	3	4	5	6	7	8	9	10	11	12
1 基礎資料整理												
2 研商諮詢會議												
3 執行細節調整												
4 資料庫維護												
5 資料庫修正												
6 資料數位化作業												
7 族語新聞轉譯												
8 太魯閣語料 2 卷及 6 則族語新聞建置完成												
9 太魯閣語料 3 卷及 6 則族語新聞建置完成												
10 提交「108 年數位族語資料庫建置報告」												

二、其他建議

(一) 未來可優先大量轉譯族語新聞，建置大型口語語料庫。

族語新聞至今已累積大量的資料，但需要大型計畫才能支持將大批新聞口語語料轉為書面語料。

(二) 培育相關口語轉譯文字人員。

未來需大量的技術人員，能將口語語料轉譯成文字，需能夠辨認語音及瞭解該族的書寫規範，需具一致性。

(三) 跨領域與教學及語言發展應用結合。

未來資料庫應擴大應用功能，如結合語言學習，並聘請專家協助建置原住民族語語料庫。

陸、參考文獻及網站

一、參考文獻

Crystal, David. 1991. *Dictionary of Linguistics and Phonetics*. 3rd edition. Cambridge, MA: Basil Blackwell.

Zeitoun, Elizabeth, Yu Ching-hua and Weng Cui-xia. 2003. The Formosan Language Archive: development of a multimedia tool to salvage the languages and oral traditions of the indigenous tribes of Taiwan. *Oceanic Linguistics*, 42.1:218-232.

大衛·克里斯托 (David Crystal) 著，蔡淑菁 (譯)。2010。《語言的秘密》(Language Death)。臺北市：貓頭鷹出版。

朱清義。2013。〈阿美語書寫系統：o 及 u 的爭議〉。《原教界》。期 49，頁 62-64。

何德華、董瑪女。2016。《雅美 (達悟) 語語法概論》。新北市：原住民族委員會。

李壬癸。2009。〈日本學者對於臺灣南島語言研究的貢獻〉。《臺灣語文研究》，4:1-20。

李壬癸。2010。《珍惜臺灣南島語言》。臺北市：前衛。

原住民族委員會。2013。原住民族語言振興第 2 期六年計畫 (103-108 年)。網址：<https://goo.gl/qs6LPe> (2017.11 瀏覽)。

陳浩然、潘依婷。2017。〈語料庫與華語教學〉。收入於陳浩然編《語料庫與華語教學》。臺北：高等教育。

程婉如。2008。拓展臺灣數位典藏計畫：淺談光學字型辨識技術 (OCR, Optical Character Recognition)。網址：
<http://content.teldap.tw/index/?p=529> (2017.12 瀏覽)。

齊莉莎、余清華。2004。〈臺灣南島語數位典藏：詞類與標記原則〉。
<http://linganchor.sinica.edu.tw/data/file/LC040907LC10.pdf>。

齊莉莎、余清華。2007。〈臺灣南島語數位典藏〉。《原住民族語言發展—理論與實務研討會》，臺北。

<http://ip194097.ntcu.edu.tw/giankiu/GTH/2007/fldc/lunbun/1-2.pdf>

齊莉莎。2013。〈臺灣南島語典藏〉。發表於《兩岸瀕危語言典藏及語言振興訓練國際研討會》，臺中。

http://ctldc.cs.pu.edu.tw/file/0720_5.pdf

原住民族委員會。2013。「原住民族語言振興第2期六年計畫」(103-108年)。網址：<https://goo.gl/qs6LPe> (2017.11 瀏覽)。

原住民族委員會。2018。「107年度原住民族瀕危語言復振補助計畫」。網址：<https://goo.gl/orrWbS> (2018.11 瀏覽)。

原住民族語言研究發展中心。2017。「原住民族語言研究發展中心數位族語資料庫建置評析報告」。網址：<https://goo.gl/MiVhr5> (2018年11 瀏覽)。

二、參考網站

AILLA (The Archive of the Indigenous Languages of Latin America)，拉丁美洲原住民族語言檔案館，網址：<https://ailla.utexas.org/>

Alilin 原住民族電子書城，網址：<http://alilin.apc.gov.tw/tw/>

ASEDA (The Aboriginal Studies Electronic Data Archive)，澳洲原住民研究電子數據檔案庫，網址 <http://aseda.aiatsis.gov.au>

Corpus of Contemporary American English，美國當代英語語料庫，網址：<https://corpus.byu.edu/>

Daly Languages (AUSTRALIA)，澳洲 Daly 語語料庫，網址：<http://www.dalylanguages.org/>

Glottolog，世界語言參考文獻網，網址：<http://glottolog.org/langdoc/langdocinformation>

Kevin Scannell on Corpus-based language technology for indigenous and minority languages，網址：

<https://www.youtube.com/watch?v=W05YmEl093o>。

LDC (Linguistic data Consortium, University of Pennsylvania)，美國賓州

大學語言資料庫，網址：<https://www ldc upenn edu/>

OLAC: Open Language Archives Community，開放語言檔案社群網，網

址：<http://www language archives org/>

PARADISEC (Pacific and Regional Archive for Digital Sources in

Endangered Cultures)，太平洋地區檔案數位資源瀕危文化庫，網

址：<http://www paradisec org au/>

UNESCO Interactive Atlas of the World's Languages in Danger 聯合國教科

文組織世界瀕危語言地圖，<http://www.unesco.org/languages->

[atlas/index.php](http://www.unesco.org/languages-atlas/index.php)

中央研究院語言學研究所「語言典藏計畫」之臺灣南島語數位典藏（目前已停用），網址：

<http://languagearchives.sinica.edu.tw/cht/index.php.html>。

臺大臺灣南島語多媒體語料庫，網址：

http://corpus.linguistics.ntu.edu.tw/index_zh.php。

臺灣原住民族數位典藏資料庫，網址：

<http://dore.tacp.gov.tw/tacp/pingpu/index.html>

臺灣聖經公會聖經網站，網址：<http://cb.fhl.net/>

臺灣語言學學會（語典及語料庫），網址：<http://linguist.tw/zh->

[tw/links/%E7%9B%B8%E9%97%9C%E7%B6%B2%E8%B7%AF%E](http://linguist.tw/zh-tw/links/%E7%9B%B8%E9%97%9C%E7%B6%B2%E8%B7%AF%E)

[8%B3%87%E6%BA%90%E9%80%A3%E7%B5%90/dictionaries%20](http://linguist.tw/zh-tw/links/%E7%9B%B8%E9%97%9C%E7%B6%B2%E8%B7%AF%E)

[and%20corpus](http://linguist.tw/zh-tw/links/%E7%9B%B8%E9%97%9C%E7%B6%B2%E8%B7%AF%E)。

花蓮縣國小本土語文領域輔導團（太魯閣字典），網址：

<http://teacher.hlc.edu.tw/imain9.asp?id=81&sid=259>

阿美族、雅美族口語傳說採錄翻譯資料數位典藏計畫，網址：

<http://aya.ioe.sinica.edu.tw/>

- 阿美語萌典，網址：<https://amis.moedict.tw/>
- 阿美語線上讀經，網址：http://www.amis100.org.tw/cudad_0_0.htm
- 政大原民中心電子書城（最新九階教材 2017-07-04 第 3 刷），網址：
<http://ebook.alcd.tw/>
- 原住民族委員會臺灣原住民族圖書資訊中心，網址：
<http://www.tiprc.org.tw/common/index.php>
- 原住民族語言研究發展中心（原住民族委員會）歷年研究成果，網址：
<http://ilrdc.tw/achievement/>
- 原住民族語言線上詞典（原住民族委員會），網址：[https://e-](https://e-dictionary.apc.gov.tw/)
[dictionary.apc.gov.tw/](https://e-dictionary.apc.gov.tw/)
- 原住民族語言線上詞典，網址：<http://e-dictionary.apc.gov.tw/Index.htm>
- 國民記憶庫—臺灣故事島，網址：<https://storytaiwan.tw/default.html>
- 國家研究院本土教育資源網，網址：
<http://3w.naer.edu.tw/natural/index.htm>
- 國家圖書館臺灣書目整合查詢系統(SMRT)，網址：
<http://metadata.ncl.edu.tw/blstkmc/blstkmc#tudorkmtop>
- 教育部朗聲四起（原住民族語朗讀文章），網址：[https://alcd-](https://alcd-con.tw/alr/lobby.php)
[con.tw/alr/lobby.php](https://alcd-con.tw/alr/lobby.php)
- 族語 E 樂園（原住民族委員會），網址：<http://web.klokah.tw/>
- 創用 CC 授權條款，網址：<http://creativecommons.tw/license>
- 達悟語線上學習網，網址：
<http://yamiproject.cs.pu.edu.tw/elearn/?&lang=zhtw>
- 數位典藏國家型計畫—臺灣社會人文電子影音數位博物館計畫，網址：
<http://proj1.sinica.edu.tw/~video/search/index1.html>
- 蘭嶼媒體與文化數位典藏，網址：<http://lanyu.nctu.edu.tw/database/>
- 蘭嶼達悟語口語資料典藏網，網址：
http://yamiproject.cs.pu.edu.tw/yami/yami_ch/ch_index_flash.htm

蘭嶼達悟語線上學習網，網址：

<http://yamiproject.cs.pu.edu.tw/elearn/search.php>

附錄一、原住民族語言研究發展中心數位族語資料庫首頁介面翻譯清單

為加強原住民族語言的使用，並提升族語能見度，本年度在建置資料庫時，亦提及首頁族語介面需求，目前仍在持續完成中，已翻譯完成之清單，相關資訊未於網路版本報告公開。

附錄二、諮詢訪談及重要會議相關

諮詢、訪談及重要會議清單，相關資訊未於網路版本報告公開。

附錄三、計畫相關會議簽到統計及會議記錄

2018 年辦理之數位族語資料庫建置、轉譯工作、修正研商會議，相關資訊未於網路版本報告公開。

