

## 壹、計畫緣起

近十餘年來隨著國內族群意識的逐漸提升與相關政策措施之推展，在官方、學界與民間產出了許多原住民族語言相關之研究、教學與語料相關資源，如字辭典、詞彙表、語法、語料庫、教材、教案、教具、繪本及近年來因應網路與雲端平台趨勢所發展之諸多網路資源，然而，因缺乏共同匯整之平台與單位，除致使各項資源散佈於各個單位、組織外，亦有內容標準不一致之狀況產生，因而在教學或語言使用上容易產生混淆之狀況。

為使各項資源得以發揮最大效益，必須建立一個資源整合之平台與管理維護之常設單位，持續透過調查、整理、審查、取得授權、建檔、發佈等程序，使相關資源得以整合提供各界應用、學習與研究。

然而語文資源之種類繁多、格式多樣、來源多元，資源整合首先要考量的將是資源規模之廣度與資源管理之深度，資源規模之大小將涉及到所需時間、人力與經費，資源管理之程度則影響資源內容標準化之狀況，以及資源整合後之可利用程度，此兩者若無明確之定義與規範，於整合時無所依歸，進而使資源之整合產生混亂與無效率。

此外，在政府資源的使用上亦應進行整合，以發揮資源最大效益，因此在考量資源有限以及人力、物力及時間之限制下，本計畫建議原住民族語文資源整合可先著重於原住民族委員會所編制之數位教材內容、線上字詞典、數位典藏與資料庫等，並採用如都柏林核心集 (Dublin Core Metadata Element Set) 這類的國際統一標準建置資源聯合目錄，使各實體與數位資源可進行彙整，並由原民會之「臺灣原住民族圖書資訊中心」計畫及該計畫已建置完成之「臺灣原住民族資訊資源網 (<http://www.tipp.org.tw/>)」統一管理，避免資源分散或重工之狀況發生，有效運用資源，使資源發揮最大之效益。

## 貳、計畫目標

1. 配合現有資源平台，彙整原住民族委員會所編制之實體或數位資源。
2. 以統一之資源彙編標準建置資源聯合目錄，整合各項資源內容，以便各項計畫管理運用。
3. 建立語文資源架構，檢視會內發展之語文資源溢缺狀態，並提出修正建議或發展創制計畫。
4. 依據書寫系統及標點符號使用原則之標準，檢視各項資源之文本狀態，並制定各項教、學資源之修訂原則與規劃，以提升其正確性與品質。

## 參、計畫內容

本計畫將先就語文資源進行定義並說明本次計畫之目標資源範圍，其次說明資源整合之程序與步驟，並提出期程規劃，最後提出相關建議。以下就各項工作進行說明：

### 一、國內原住民族語文資源定義與概況

近十年來，隨著國內民族意識的復甦以及相關政策的推動，從中央到地方均有諸多有關原住民族語相關資源的彙編與產生，而隨著多媒體技術與網路科技的發展，近幾年來族語相關教育資源亦有影音化、數位化與雲端化之趨勢。而在了解語文資源概況時，可從「來源」、「類別」與「型態」等三個面向進行探究，不同的面向將涉及語文資源整合時所需考量的層面。

資源的來源可分為官方、學界與民間等三類，官方資源為政府單位相關計畫所產出的語文資源成果，通常有字辭典、教材彙編或研究報告等類別；學界資源則為學術界經過田野調查或研究所產生，如期刊論文、語料庫、田調之影音或文字紀錄；民間資源則最為多元，諸如個人創作、紀錄，族語教師自行發展之教材、教案、教具，組織團體發展之文本資料，如聖經、其他文字或影音出版品等等。

語文資源之來源會涉及到整合資源時所採用的方式，政府資源與學術資源可透過單位協調的方式取得相關資料，然民間資源則可能將涉及智慧財產權授權之法律程序以及授權費或購置資源等相關經費使用。本次計畫雖僅針對原住民族委員會所編制之語文資源進行整合，應無所謂智慧財產權取得之問題，唯整合時仍須徹底檢視各項資源是否確實擁有智慧財產權，以避免可能之相關問題，另本計畫將提出整合國內現有之原住民族語料庫之後續建議，此部份除須考量資料與系統統整性外，則可能會遇到上述之智財授權之問題。

表 1：語文資源類別形態表

類別	形態			
	紙本	電子檔 (文字、聲音、影像)	網路	實體 (如教具)
<b>1. 語料</b>				
1.1 字詞典	✓	✓	✓	
1.2 田調資料	✓	✓	✓	
1.3 語料庫		✓	✓	
1.4 創作與紀錄		✓		✓
1.5 試題	✓	✓		
<b>2. 學術資源</b>				
2.1 論文	✓	✓	✓	
2.2 期刊	✓	✓	✓	
2.3 研究報告	✓	✓	✓	
<b>3. 教學資源</b>				
3.1 教材	✓	✓	✓	
3.2 教案	✓	✓	✓	
3.3 教具	✓	✓		✓
3.4 人才資料庫		✓		

### 二、資源整合範圍

上述各項語文資源種類繁多，來源及形態不一，如欲進行全面性的整合則需要長期之大型計劃，所需經費及人力龐大，故本次計畫將資源整合範圍聚焦於原

住民族委員會所編制、出版之各項語文資源，包含實體與數位之教材、字辭典、詞彙表、數位典藏或資料庫等，先建立資源整合架構，未來再逐步推展至民間與學界之資源。

## 肆、整合步驟

### 一、語文資源調查

語文資源整合的首要工作乃是調查現有資源有哪些，因本次計畫之資源範圍限定在原民會所出版之語文資源，故可先召開會內協調會，說明資源整合計畫，並提供語文資源線上調查表提供各單位填寫以統整資源概況，調查表內需有單位名稱、填表人資訊、資源名稱、形態（實體、數位或網路）、所在位置、所有權、管理權、存續狀況（如彙編中、已完成、待修正…等）等相關欄位，以便了解各項資源歸屬情形與可處理方式，調查表草案如下：

表 2：語文資源調查表（草案）

原住民族委員會 語文資源調查表			
單位名稱			
填表人		連絡電話	
email			
資源名稱		資源類型 (實體、數位、網路、其他)	
資源數量			
資源位置			
所有權歸屬			
管理權歸屬			
存續狀況 (如彙編中、已完成、待修正…等)			
備註			

## 二、資源篩選

彙整各單位資源調查表後可了解會內語文資源概況，然語文資源之存續狀況會影響資源是否適合進行整合運用，因整合後資源可能將提供教學或研究使用，因此資源之正確性與時效性便顯得十分重要，以教材為例，九階教材自 94 年編輯完成後已有多年未曾修訂，其內容正確性與後來編制之補充教材或千詞表、字母篇、生活會話篇之內容一致性亦有若干差異，倘若一併整合公告，將造成使用者之混亂。

因此進行整合前應就資源狀況進行分析、篩選與分類，部分資源已知有後續修訂計畫者，應特別標註更新期限，網路資源若無後續維護計畫者，亦應特別註明，並於後續整合工作時將資源內容併入資源整合平台，避免因原始資源網站消失而遺失。

## 三、建立聯合目錄標準

資源整合中最常遇到的問題為各項資源目錄編碼不一，導致搜尋困難，因此應以統一資料元素標準建立聯合目錄，以作為資源整合之基礎，如都柏林核心集 (Dublin Core Metadata Element Set, 簡稱 DCMES 或 DC)。

都柏林核心集是由都柏林後設資料組織 (The Dublin Core Metadata Initiative, DCMI) 所發展，為一組具簡單性、普遍性、延展性、國際性的後設資料元素，用以描述廣泛的網路資源，作為跨學科領域資源分享的交換格式。透過都柏林核心集的對照，不同典藏機構可以達到資訊彼此交換、共享之目標。都柏林核心集具有三大原則<sup>1</sup>：

<sup>1</sup> 摘自「都柏林核心集最佳實務指引」，陳亞寧等作者群，後設資料工作組，數位典藏國家型科技計畫，2010 年 10 月。

#### (一) 一對一原則 (The One-to-One Principle)

每一筆都柏林核心集の後設資料紀錄僅能描述或對應一個資源的一個形式或版本，而非擴展至該資源的其他形式，如同一件教材具有實體與數位版本，則應視為兩筆資料，實體及數位版本皆有各自的後設資料時，則兩筆後設資料之間可使用關聯元素說明兩者之關聯，讓使用者知道此資源的另一個格式。

#### (二) 簡化原則 (The Dumb-down Principle)

此原則係指每個元素之下，使用者可忽略任何的限定詞，直接使用未限定的資料值。刪去限定詞後所保留下來的資料值，雖然可能會失去一些特定性，對資源發掘仍是正確而且有用。

#### (三) 適宜的資料值 (Appropriate values)

特定元素或限定詞的最佳實務 (best practice) 會隨著脈絡而有所不同，一般而言，後設資料解譯者總是機器，結果是實作者所無法預測。因此最佳實務可能對如何建構 DC 後設資料加諸特定限制，但對於發掘資源的需求須加以留意。

### 四、 建構或選擇語文資源整合平台

依據統一資料元素標準建立聯合目錄後，即可將篩選後之資源進行授權、整合、轉移或複製等工作，但整合後之資源應有一平台供查詢使用，本計畫建議利用原民會已建置之「臺灣原住民族圖書資訊中心」及「臺灣原住民族資訊資源網 (<http://www.tipp.org.tw/>)」作為資源整合平台，除可避免重複建置外，亦可有效利用既有資源，並可避免使用者之混淆。

### 五、 資源管理與維護

前述之資源整合可分為如書籍、教育等實體資源與數位檔案、網路資源等非實體資源，實體資源之移動與重製較為容易，然非實體資源則涉及技術性問題，如原住民族語 E 學園上除了檔案下載外，另有線上測驗之功能，若要整合於同一平台，則在功能設置上恐需進行大規模之網站改寫，因此建議採連結方式進行網路資源整合，以增進整合之時效性，但未來可考量發展長期計畫，避免設置類似功能之諸多網站，造成使用者之混淆。

再者，語文資料會面臨資料更新之問題，此部分特別注意各資料之更新提醒註記，負責整合語文資源的單位並需與各資源來源單位保持密切聯繫，定期更新資源現況，以保持語文資源之正確性與時效性。

此外，有關正確性的部分另須注意的是有關原住民族語言書寫系統及標點符號使用原則之問題，原住民族語言書寫系統已頒布十年，在使用上應有需要調整或修正之處，而隨著族語書寫之普及，目前面臨的有關借詞、新詞或外文(如英文、德文、中文)與族語書寫同時出現時之格式、用法、規則應如何處理，亦為有待解決之問題，而此類問題涉及的是語文資源文本內容的正確性與不同資源間的一致性，需有長期計畫進行檢視修正，並建立修正規則，針對後續各項語文資源計畫進行統一之調整與修正。

## 伍、 結論與建議

### 一、 結論

完整之語文資源整合計畫涉及所欲整合之「資源規模」、「資源分布」與「資源加工」等三大面向，本計畫僅就原住民族委員會之內部資源進行整合，以建立整合架構與標準流程，未來可依據經費預算規劃整合民間與學界資源，對於國內之原住民族語言之教育、學習與研究等領域將有長足之助益。完整之語文資源整合計畫架構圖草擬如下：

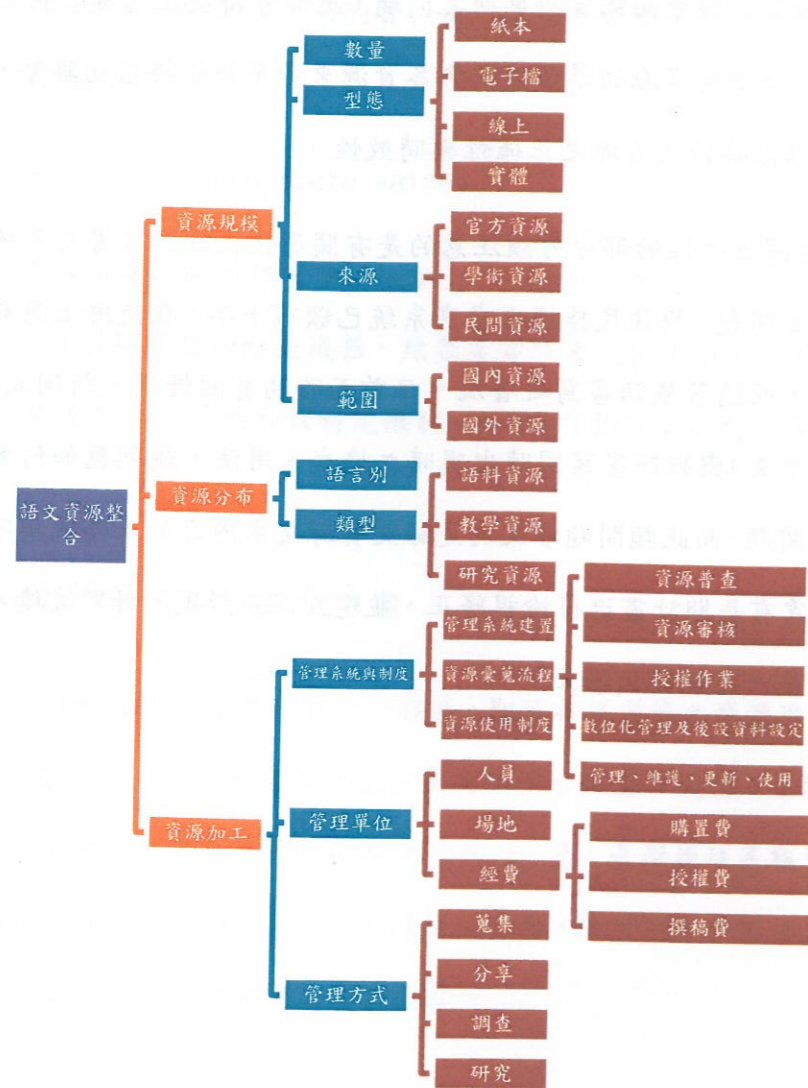


圖 1：語文資源整合架構圖

### 二、 建議

#### (一) 標準修訂

本計畫僅就資源本身進行調查與整合，並未就「資源內容」進行校正，建議研擬資源校正計畫，先就原住民族語言書寫系統、標點符號使用原則目前所面臨之問題提出調查報告，並設立專案進行修訂研究後，依據修訂後之書寫系統與標點符號使用原則，制定各項教、學資源之修訂原則與規劃，以提升其正確性與品質。

#### (二) 資源檢視

建議可透過資源彙整之過程，結合語發中心之教學研究計畫與「臺灣原住民族圖書資訊中心」之現有成果，檢視教學或研究資源之缺口，並建立資源發展計畫補足所需資源，以使整體之教、學與研究資源系統更加完整。

#### (三) 研究發展

本次計畫僅針對原民會所編製出版之語文資源進行整合，建議後續計畫可統整國內如中研院「台灣南島語數位典藏」、台灣大學「台灣南語多媒體語料庫」、靜宜大學「蘭嶼達悟語口語資料典藏網」等現有語料資源，建置官方語料庫長期發展計畫，作為台灣原住民族語言研究之重要基礎建設。