

# 財團法人原住民族語言研究發展基金會 112 年度預算案 審查報告（修正本）

中華民國 112 年 5 月 3 日  
華總一經字第 11200035351 號

(一)工作計畫：應依據收入與支出審查結果，隨同調整。

(二)收入、支出及餘絀：

- 1.收入：1 億 7,762 萬元，照列。
- 2.支出：1 億 7,760 萬元，照列。
- 3.本期賸餘：2 萬元，照列。

(三)通過決議 2 項：

- 1.原住民族語言研究發展基金會自 108 年成立時，即把原住民族語言人工智慧系統建置計畫納入業務職掌，惟原住民族口語語料之採集，109 年總計僅 36 小時，110 年總計僅 82 小時，部分族群如噶瑪蘭族兩個年度更分別只有 15 分鐘；109 至 111 年所採集之口語語料，目前經校訂者只占總數之 23.96%，校訂速度跟不上採集速度。

以澳大利亞研究委員會-語言動力學研究中心（ARC CoEDL）案例，CoEDL 與原住民社區合作收集的 40,000 小時錄音，為了解決人類語言學家需要將近 200 萬小時來轉錄（語音轉錄為文字）和處理的問題，研究人員 Ben Foley 和 Janet Wiles 教授一直在研究人工智慧技術以尋找提高效率的方法。CoEDL 和谷歌建立合作關係，提供了一個名為 TensorFlow 的開源人工智慧平台，可以使用它來構建自己的人工智慧和機器學習系統，以轉錄原住民族語言並構建可以支持語言工作的人工智慧模型。到目前為止（2018），他們已經為 6 種澳洲原住民族語言以及另外 5 種在亞太地區使用的語言建立了初始模型。

故請原住民族語言研究發展基金會參考各國案例，擬定原住民族口語語料採集及校訂提升方案及未來年度關鍵績效指標，以利語言保存及傳承，以及原住民族語言人工智慧系統建置。

原住民族語言口語語料採集所累積之時數較少，採集後的口語資料校訂效率亦較慢，不利於原住民族語言快速流失下的語言保存和傳承，爰請於 2 個月內擬定原住民族口語語料採集及校訂提升方案及未來年度關鍵績效指標，向立法院內政委員會及提案委員提出書面報告。

2. 為建構族語專才資料庫，原住民族語言研究發展基金會設有「族語人才資料庫」、「族語認證資料庫」。近期有族語專門人才反應，接到有關族語相關學術調查之電話訪談與 email 問卷，恐是資料庫資料外洩所致。據瞭解，目前查詢族語人才紀錄，僅記錄查詢者上線時間登載，並未完整記錄查詢者所查詢之內容，資料控管明顯有不足之處，且原住民族語言研究發展基金會針對研究計畫申請案件，也未設有倫理審查機制。爰此，要求應重新檢討資料查詢機制，查詢者網路查詢路徑、查詢目的及所有查詢資料，應做完全記錄。針對研究需要申請查詢資料，應設有倫理審查機制，以確保個人資料妥善維護。