

學術調查研究資料庫的變革與創新

廖培珊*

為保存國內調查資料並加以推廣利用，中央研究院人文社會科學研究中心所屬之調查研究專題中心（簡稱調研中心），規劃設置了「學術調查研究資料庫」（簡稱資料庫），以蒐集、整理、彙整、釋出學術調查資料原始檔案為工作任務，收藏涵蓋社會、政治、經濟、教育、心理、公共衛生等研究領域，以問卷調查方式蒐集的原始資料檔案及相關參考文件，經過整理、檢誤並統一格式後，開放予國內外學術界人士申請使用。自 1994 年 11 月建置資料庫以來，陸續蒐羅了各種政府調查資料及國內學術調查資料，這些調查資料的來源有三種：一是以簽訂授權合約方式取得國內主要提供研究經費之政府機關補助的研究計畫，或是由政府機關自行辦理調查的調查資料；二是調研中心協辦調查業務之研究計畫；三為學者或國內各研究單位捐贈之資料。

為了讓研究者能夠更為便捷且有效率的獲取所需要的調查資料，資料庫於 2003 年起以會員制的方式，透過網際網路的各項功能，讓會員可隨時查詢及下載資料使用，若有疑問還可以透過專人服務進行線上諮詢，提供了更全面性的服務（參見王文心，2003）。¹ 同時，藉由科技部及其前身國家科學委員會多年來規劃推動補助計畫的資助，除了持續改進資料服務的功能，豐富資料庫內容，並定期發行「學術調查研究資料庫」最新消息電子報及通訊等刊物外，近年來亦以多種方式推廣、提升調查資料品質，以及進行調查資料的加值與再利用。

為了讓研究者可以更迅速精確的搜尋所需要的資料，持續建置概念詞庫，在推廣資料使用的工作上，不僅將資料庫的網站以嶄新面貌呈現，更規劃建置了英文網站，並陸續將調查資料與相關文件英文化，以方便更多的國內外學者使用。此外，亦持續於 Nesstar 軟體上建置後設資料，並提供線上分析功能，以提升使用者的使用便利性。有鑑於調查資料日趨受到重視，資料庫參考國外大型資料中心的作法，引進了數位物件辨識碼（Digital Object

* 中央研究院人文社會科學研究中心調查研究專題中心副研究員

¹ 學術調查研究資料庫的網址為：<https://srda.sinica.edu.tw/>



Identifier, DOI), 方便研究者標示資料來源, 提升資料檔與資料庫的能見度。以下說明資料庫近年來的發展與規劃。

一、增強資料庫的服務與功能

為了提供會員及學術界更為完善的資料服務與諮詢工作, 資料庫的新網頁已經自 2011 年 5 月起正式對外公開使用, 並且持續進行各項功能的加強。近幾年來較為主要的變革包含「學術調查研究資料庫」會員制的改變、釋出資料檔及相關使用說明的重整與更新、線上影音專區之設立, 以及限制性資料的規劃與使用等。

(一)「學術調查研究資料庫」會員制

為了便利學術界取用調查資料, 並保障資料庫與使用者雙方的權利與義務, 調研中心自 2003 年 5 月開始推動會員制, 凡符合「『學術調查研究資料庫』會員約定條款」會員資格並同意遵守約定的人士, 皆可以申請成為會員。後於 2011 年 5 月修訂會員類別與會員約定條款, 原一般會員修改為一般會員(二年期), 臨時會員修改為一般會員(一年期), 並將中央研究院專任研究人員從一般會員獨立為院內會員。另外新增網路會員, 網路會員為不限資格之個人, 且網路會員帳號為永久有效。會員資格及權益均於「會員約定條款」中述明, 可至資料庫網頁查詢, 並可於線上提出申請加入會員。

針對一般會員資格即將到期的會員, 調研中心會主動與會員聯絡資料校正及資格延長事宜, 提醒會員提供最新的在職/在學證明文件影本, 以便延長其使用期限, 避免會員欲登入網站查詢資料時才發現過期, 造成使用上的不便。至於中央研究院的院內會員, 調研中心於每年 11 月以電子郵件校正資料, 資料確認無誤後, 便會主動延長使用期限。若一般會員與院內會員的帳號過期或失效後, 將自動轉為網路會員。

(二)釋出資料檔及相關文件的重整與更新

由於資料庫所蒐藏的資料檔案來源不同, 伴隨著而來的說明文件與內容也會有所差異, 造成每一筆資料檔的後設資料內容與形式也不盡相同。同時, 資訊科技的變革也使得既存的資料或文件檔案形式可能已過時, 或不符合使用者需求。因此資料庫針對需求較高的政府調查及大型學術調查資料進行清查、整理與檢誤, 然後上傳更新檔案, 以使得資料庫所釋出的資料檔內容與相關文件說明能以標準的形式呈現。

重整工作順序與內容規劃依以下順序進行資料重整：

1. 前置資料合併：合併程式、重覆樣本編號的確認、資料併檔。
2. 修改資料：精簡變項說明及選項數值說明、修改變項名稱依題號命名、複選題過錄方式的統一、新增基本問項等。
3. 檢查不合理值：轉出新的原始資料及 SPSS 程式、不合理值檢查與修改、轉檔案格式、檢誤報告重整。
4. 釋出準備：刪除不釋出的變項、整理相關說明文件。

重整後的各項資料檔及說明文件亦同時上傳 Nesstar 線上分析系統，並利用網頁、電子報等各項管道公告，以利資料使用者獲得最即時的資訊。

(三) 線上影音專區之設立

為了推廣可提升調查資料品質的方法，調研中心多年來辦理「調查研究資料整理與檢誤研習班」，以科技部補助之計畫主持人或其計畫助理為主要研習對象，目的在於增強參與學員熟練整理調查資料的技巧，期能對調查資料品質有正面的幫助。為節省舉辦研習班的人力與時間等成本，以及服務遠程或時間無法配合的計畫主持人或其計畫助理，自 2013 年度暫不舉辦「調查研究資料整理與檢誤研習班」，而改以網路影片教學的方式，提供整理調查資料的技巧，讓有興趣的資料使用者可隨時上線觀看，期能對調查資料品質有正面的幫助。課程內容包括：學術調查資料庫資料簡介、資料整理程序、編碼設計概論、資料整理與檢核之實務(含 SPSS、STATA 上機實例練習)等。

此外，由於「學術調查研究資料庫」已有數量龐大與豐富的資料蒐藏，除了持續爭取更多的資料入庫之外，次級資料的推廣更是迫切的工作重點，近年來調研中心已舉辦多場資料使用或分析工作坊，邀請計畫主持人或相關領域學者，介紹資料特色、分享分析經驗，讓學員能對資料有更深入一層的了解，以推廣次級調查資料的使用，並提升學術研究能力。在授課人員的同意下，亦將多場工作坊的內容上傳至影音專區，希望以此提供更專業與便利的服務平台。影音專區可於資料庫網頁的「相關資源」中觀賞使用。

(四) 限制性資料的規劃與使用

調查資料中有時會包含隱私或機敏資料，可用以辨識特定的個人、班級、學校、村里等。為了降低受訪者被辨識出的風險，並符合《個人資料保護法》的要求，部分調查資料除了一般會員可於網頁上下載使用的版本外，也規劃了需要另行申請的限制版資料，並在有所管制的情況下使用。通過審核的申請者，必須利用調研中心的限制性資料使用室，以符合使用規則的方式進行資料分析，事後可攜出不含任何足資識別個別樣本 / 單位之資料之統



計分析結果與程式檔。

除了現場使用外，並於 2011 年規劃啟用了限制性資料的遠距服務 (Remote Service) 功能，提供更為便利的服務。使用者必須先申請帳號及密碼，並且有固定 IP 位址，才可以在有效期間內連線使用遠距服務，並需遵守使用規則來使用限制性資料。為了確保資料庫中所保管之各項資料檔的安全性，資料庫引進了 ISO 27001:2005 資訊安全管理系統，並於 2010 年通過驗證，爾後亦持續取得認證。目前的限制性資料已包含十個大型調查計畫所蒐集之資料，藏量與涵蓋領域也持續增加中。

二、提升調查資料的價值與使用便利性

除了持續加強資料庫的服務功能與資料蒐藏內容，提升調查資料的價值，並連帶的提高使用上的便利性，亦是資料庫持續努力的目標。以下僅列舉三項較具突破性的重要工作，一是持續變項層級概念詞庫的建置，二是資料庫英文化的工作，三是引進數位物件辨識碼 (Digital Object Identifier, DOI)。

(一) 概念詞庫建置

過去的資料搜尋條件，多以計畫名稱、主持人、摘要、關鍵字或題目等為主，若是想搜尋未出現在問卷題目中的概念詞，常徒勞而返。因此「學術調查研究資料庫」由大型學術調查資料開始著手，分析問卷題目中的概念，並規劃建置變項層級的中、英文概念詞，並將既有成果逐步擴展到其他相關學門領域的調查計畫。目前概念詞庫搭配資料庫「以題項內容查詢」的搜尋功能，已成為研究者找尋資料的利器；資料搜尋的範圍同時包括政府調查資料與學術調查資料，並整合了資料檔相關資訊的瀏覽與線上分析的功能，以提高搜尋的便捷與效益。

概念詞庫的建置是以該調查計畫主持人為諮詢對象，取得原始問卷設計時的完整概念，以獲得最精準的分析結果。由於類似的題目可能出現在不同的問卷，不同領域的專家學者提供的概念詞亦可能有所不同，因此需要彙整這些相似題目的概念詞，進行一致性的作業，並更新已建置的概念詞內容。截至 2013 年年底為止，已完成「臺灣社會意向調查」44 份調查資料、「臺灣基因體意向計畫」12 份調查資料、「臺灣教育長期追蹤資料庫」10 份調查資料，以及 6 份科技部個別型計畫的概念詞建置，並匯入資料庫供研究者查詢使用，隨著詞庫越來越豐富，相信對研究者的幫助也會越大。

(二) 資料庫英文化

英文化的工作分成兩部分，一是建置學術調查研究資料庫的英文版網站，二是資料檔英文化。依據本資料庫的規劃，這兩項工作將同時帶給國內外研究者極大的利益與便利：國外使用者不但可從英文網站獲得與中文網站等量的資訊與便利性，也能搜尋得到英文版的資料檔；而當國內使用者利用資料庫「以題項內容查詢」的搜尋方式時，即使以中文搜尋資料時，也可直接選擇是否使用英文版資料，減少自行翻譯資料內容的困擾。資料庫網站的英文化，歷經先期的規劃與系統分析，與廠商不斷的討論、修改、測試與調整，目前已對外正式開放。

而資料檔英文化部分，由於資料庫所蒐藏的資料量十分龐大，英文化的工作需妥善規劃，並選擇研究價值性較高且使用率多的資料先處理。起始的工作先從「臺灣社會變遷基本調查」及「臺灣華人家庭動態」計畫著手，雖然計畫早已具備國際視野，先一步製作英文版問卷，但是從英文問卷到英文資料檔釋出，還有許多工作要處理，除了要先製作欄位定義程式，進而產生資料檔、轉換各種統計軟體系統檔，並製作英文版資料使用說明，最後經過複檢所有的釋出檔案與文件，英文版資料才得以釋出，目前為止完成英文化作業的資料檔，已經隨同英文網站的正式上線，陸續對外提供下載服務。

(三) 數位物件辨識碼

在撰寫論文時，對於次級資料分析的資料檔之說明日趨受到重視，這些資料檔如果可以得到使用者正確的引用，就能推廣研究成果，鼓勵研究人員蒐集高品質的調查資料，並得到應有的榮譽；另一方面，對於資料釋出單位，也能提高其能見度，並達到推廣次級資料的目標。目前國外相關資料庫都正在推動調查資料的引用，而數位物件辨識碼 (Digital Object Identifier, DOI) 是目前最廣為採用的標準。

數位物件識別碼，是物件在網路上的唯一識別碼，可以提供穩定、持續、可信的連結服務。「學術調查研究資料庫」所提供的每一筆網路下載資料，即可視為一個物件，因此每一筆資料都可以註冊一個 DOI 編號，當資料使用者在發表著作時，於資料引用說明或是參考文獻中列出該筆資料對應的 DOI，其他人就可以方便的連結到資料庫，取得相同的資料檔。同時 DOI 為一國際標準及技術，可與市場接軌，提升資料庫的國際能見度及影響力。以目前 SRDA 擁有的眾多釋出資料而言，除了彙整相關資訊，規劃 DOI 編碼原則，並提供 DOI 授權單位進行註冊外，資料庫也對應建置相關欄位與說明，



在投入相當人力與財力之下，今年已經完成一千多筆資料的 DOI 建置，新增網頁的 DOI 欄位。有關引用 DOI 於參考書目的寫作格式，可參見《人文與社會科學簡訊》15 卷 1 期文章的介紹（邱亦秀、廖培珊，2013）。

三、資料庫會員滿意度與未來規劃

由於近幾年資料庫已多有變革，並新增許多服務功能與平台，因此資料庫在 2013 年底，就會員滿意度進行網路問卷調查，希望藉由此次調查，了解會員們對資料庫之各項服務項目滿意度並提出改善建議，作為日後資料庫改進與努力的方向。綜合此次調查結果，資料庫的整體表現獲得受訪會員的高度肯定，包含資料取得的方便性、資料品質、資料蒐藏數量、工作人員專業能力、資料庫網頁使用的便利性等。但是，對於新近增加的服務功能，尤其是影音專區的教學影片，則有五成以上的會員不甚了解；會員們也建議蒐藏資料所涵蓋的領域需要繼續擴充，顯見此為資料庫需要更努力爭取與宣傳的部分。

設置「學術調查研究資料庫」有助於調查研究方法的交流與進步，國內外皆然。我們除不斷的充實服務內容與提升資料使用品質之外，也亟盼國內學者能認同資料共享的觀念並以實際行動支持，進而帶動調查方法學術研究的蓬勃發展。目前資料庫目標朝向國際化發展，不但建立英文網站，並製作資料的英文版，逐步落實資料庫的英文化。除了持續進行前述的各項重要工作外，亦規劃進行追蹤、多波次、或對象多元之大型調查資料的加值工作，以製作合併檔案的方式，讓使用者可以在平台上進行跨樣本或是跨波次的分析。由於各個大型調查的資料釋出經驗不一，而且目前個資保護的議題日益受到重視，因此資料庫會根據累積的經驗，提供計畫關於資料釋出規劃的建議，利用限制性資料釋出的機制（遠距服務、現場使用），提供安全的釋出環境，爭取更豐富的資料內容，讓使用者可以進行更細緻、更多元的分析。

參考文獻

- 王文心（2003）。〈『學術調查研究資料庫』蒐藏資料簡介及會員制度說明〉。《人文與社會科學簡訊》，第 5 卷第 1 期，頁 48-51。
- 邱亦秀、廖培珊（2013）。〈您引用了嗎？以 SRDA 為例淺談調查數據資料的參考書目格式〉。《人文與社會科學簡訊》，第 15 卷第 1 期，頁 131-135。