

SRDA 學術調查研究資料庫通訊

季刊 第21期

Survey Research Data Archive Newsletter

本期目錄

■ 最新消息	1
■ 資料釋出消息	4
■ 研究計畫調查資料管理.....	6
■ 台灣教育長期追蹤資料庫——資料使用常見問題集.....	14
■ 資料使用狀況	20
■ 會員申請程序說明.....	21
■ 捐贈資料說明	22
■ 徵稿啓事	23

中央研究院 人文社會科學研究中心 調查研究專題中心

中華民國九十六年六月



國內
郵資已付

臺灣郵政台北誌第395號執照登記為雜誌交寄

台北郵局許可證
台北字第2111號



最新消息

王文心

炎炎夏日何處去，調查研究專題中心舉辦一連串精彩的學術活動，歡迎您來充電！各項活動免費參加，一律採用「網路報名」，報名及最新活動訊息，請參閱 <http://www.sinica.edu.tw/as/survey/>。

■ 調查研究方法工作坊

主題：MPlus 結果解讀：從因素分析到結構方程模型

上課地點：中央研究院人社中心一樓會議室

上課時間：96年9月8日（六）09:00-12:20

名額限制：55人

報名資格：(1)大專院校教師或博士班學生

(2)有使用 MPlus 且熟悉結構方程模型

聯絡方式：楊景閔先生，電話：(02)2788-4188#205；email: minna@gate.sinica.edu.tw

■ 第七屆「調查方法與應用」學術研討會

研討會時間：民國96年9月7日（星期五）

研討會地點：中央研究院學術活動中心二樓第一會議室

報名注意事項

※報名截止日期：民國96年8月17日下午17:00前

※報名資料填寫不全及逾期報名者，恕不受理，敬請見諒，懇請配合。

※由於場地座位有限，若報名人數過多，將以現職與「調查研究」相關者優先參加。

※聯絡方式：謝芮桓小姐，電話：(02)2788-4188#506；email:srcsr@gate.sinica.edu.tw

■ 暫訂議程

8:30 ~ 9:00	報到
9:00 ~ 9:05	開幕 傅祖壇教授（中研院人社中心調查研究專題中心 執行長）
場次一：不完整資料	
主持人：李隆安	

時間	發表人	論文題目	評論人
9:05 ~ 9:30	林辰彥、江振東 傅祖壇、陳珍信	Evaluating Customers' Willingness to Pay a Premium for Non-Genetically-Modified Foods - A Case Study of GM fed Salmon	薛慧敏
9:30 ~ 09:55	王鴻龍、楊孟麗 林定香、陳俊如	不完整資料在因素分析的處理方法之研究	許玉雪
09:55 ~ 10:20	于若蓉、李隆安	追蹤調查缺漏資料的插補：以家庭動態調查為例	雷淑儀
10:20 ~ 10:45	廖培珊、杜素豪 李隆安	無作答資料之插補—以敏感題目為例	歐士田
10:45 ~ 11:00 茶 敘			
場次二：分析方法與應用			
主持人：傅祖壇			
時間	發表人	論文題目	評論人
11:00 ~ 11:25	陳信木、林佳瑩 陳雅琪	從橫斷面調查資料估計事件轉移率的方法——以婚姻表編製為例	陳玉華
11:25 ~ 11:50	江振東	「至多選取 k 項」的複選題分析	黃冠華
11:50 ~ 12:15	蔡良庭、楊志堅	評估取樣權重於檢定等級反應問卷之試題差異正確性研究	程爾觀
12:15 ~ 12:40	陳端容、溫在弘	多層次模型與地理加權迴歸之整合分析：以肥胖症為例	林季平
12:40 ~ 13:30 午 餐			
場次三：網路與調查研究			
主持人：黃 紀			
時間	發表人	論文題目	評論人
13:30 ~ 13:55	胡啓有、鄭雅雯 林原宏	電腦化機率問題解題規則測驗之系統研發及其實證研究	張源俊
13:55 ~ 14:20	林佳瑩、陳信木 郭蕙如	寬頻上網的普及化對於市內電話號碼抽樣的影響	謝邦昌

14:20 ~ 14:45	余民寧、李仁豪	網路化問卷可行嗎？—問卷長短、議題熟悉度、問卷內容的衝擊	于若蓉
14:45 ~ 15:10	彭森明、楊孟麗	網路調查設計與無回應者之調整	胡克威
15:10 ~ 15:25 茶敘			
場次四：量表信效度與訪答效應			
主持人：朱瑞玲			
時間	發表人	論文題目	評論人
15:25 ~ 15:50	李玲玲、彭少貞 何清治、徐祥明	中文運動自我效能量表信效度初探	齊力
15:50 ~ 16:15	杜素豪	探索態度量表回答模式的訪員效應	田芳華
16:15 ~ 16:40	李皇照、方正璽 闕韶儀	FRL 量表應用於台灣市場調查之探討	樓永堅
16:40 ~ 16:55 茶敘			
場次五：母體與抽樣			
主持人：洪永泰			
時間	發表人	論文題目	評論人
16:55 ~ 17:20	劉千嘉、林季平	台灣原住民母體變化於執行調查時之意涵：2000年普查資料與2007年戶政資料之應用	陳信木
17:20 ~ 17:45	蘇建州	電話調查訪問時段、加權估計對樣本結構與調查結果影響之初探：以2006年台北市長選舉TVBS民調為例	洪永泰
17:45 ~ 18:10	侯佩君	電話調查的「類拒訪」分析：「鴿派」拒訪圖像的解析	游清鑫
18:10	會議結束		



「學術調查研究資料庫」於 96 年 3 月至 5 月間開放釋出及改版訊息如下：

【政府抽樣調查資料】

■ 主計處 95 年「人力資源調查」、「人力運用調查」資料開放

人力資源調查統計，為現代國家配合經濟、社會發展不可或缺之重要統計指標，可供瞭解民間人力供應情形、勞動力就業狀況以及人力發展趨勢。調查之地區範圍為臺灣地區，包括臺灣省、臺北市及高雄市，而以現住此一地區內之普通住戶與共同事業戶內年滿十五歲，自由從事經濟活動之民間人口為對象，但不包括武裝勞動力及監管人口。

調查兼採派員面訪與電話訪問方式，由臺北市、高雄市暨臺灣省各縣市政府遴選調查員擔任。抽樣母體以最近一年「台灣地區年終戶籍村里別統計資料檔」為抽樣母體，採用「分層二段隨機抽樣法」，第一段樣本單位定為村里，而第二段樣本單位定為戶。

此外，為應各機關人力資源運用與經社決策施政需要，獲取較為詳細資料，乃於調查期間（每年 5 月）利用同一樣本附帶舉辦人力運用調查，旨在明瞭台灣地區勞動力運用、移轉及就業、失業狀況等短期變動情勢，供為訂定人力政策、推動職業訓練、改善企業經營等決策之參據。

開放資料的項目包含：資料檔、SAS 以及 SPSS 程式、問卷、過錄編碼簿以及資料整理報告。

■ 主計處 95 年「事業人力僱用狀況調查」資料開放

主計處為配合社會經濟發展情勢及人力決策之需要，定期蒐集具有前瞻性及政策性資料，並充實按月受雇員工薪資調查之不足，遂每年辦理事業人力僱用狀況調查，以深入了解產業界當期僱用人力狀況及企業相關人力運用計畫，俾供政策擬定各項經濟及勞動政策，以及提供企業經營、改善營運與人力規劃參考。

本項調查配合當時社經狀況及政策需要進行機動問項調整。95 年問卷內容為：（一）各職類短缺員工概況及僱用條件（二）公司運用派遣人力的現況、未來需要及滿意程度。

開放項目包括：問卷、資料檔、過錄編碼簿、次數分配表、欄位定義程式。

■ 主計處「人力資源擬一追蹤調查」資料開放

主計處進行之人力資源調查，為我國每月公布之人力指標之依據，更為勞動相

關研究應用之重要政府調查資料，自民國 67 年迄今已有超過 20 年的歷史。由於人力資源調查採用樣本輪換的方式進行，按照其樣本輪換規則，每期與前一期皆有一半的樣本重複，按此特性進行串連，即可將原先橫剖面資料的性質擴充為具有兩個至三個時點的短期追蹤資料，為過去 20 年間長時間序列勞動力相關研究提供重要而成本低廉的分析資料，大大提高其利用價值。

開放資料的年度為 93-95 年（年資料、月資料），以及 70-89 年（月資料）；釋出項目則包含：資料檔、SAS 程式、問卷、過錄編碼簿以及資料整理報告。

申請政府抽樣調查資料請至 <http://srda.sinica.edu.tw/govdb>，依步驟填妥申請表格及研究大綱，我們會在進行審核之後以 E-mail 通知領取資料。

【大型學術調查】

■ 「台灣教育長期追蹤資料庫」公共使用版資料改版

「台灣教育長期追蹤資料庫」(Taiwan Education Panel Survey, TEPS) 係由中央研究院、教育部、**教育研究院籌備處**和國科會共同規劃之全國性、長期調查計畫。本資料庫是以問卷調查方式，向國中、高中、高職及五專學生收集資料，再以這些學生為核心，擴及瞭解影響學生學習經驗的幾個最主要因素：如學生家長、老師等。本次主要更新第一、二波公共使用版的高中職/專科學生與老師的資料，改版內容說明於下：

w1 高中職/專科學生 v4.0：新增一般與數學分析能力變項-- w1cf3pc、w1cf3pcs、w1m3p29c、w1m3p29cs (可供第一、二波比較)；更新學生樣本權數-- w1stwt1

w1 高中職/專科學生 綜合分析能力(testing) v2.0：更新學生樣本權數-- w1stwt1a、w1stwt1b

w1 高中職/專科老師 以學生為單位 v4.0：更新學生樣本權數-- w1stwt1

w1 高中職/專科老師 以班級為單位 v4.0：更新權數-- w1cwt1；刪除三筆無效資料

w2 高中職/專科學生 v3.0：更新第二波數學分析能力變項-- w2math32、w2m1p32、w2m1p32s、w2m3p32、w2m3p32s (以更嚴謹的方式)；更新學生樣本權數-- w2stwt1；新增一般與數學分析能力變項-- w2cf3pc、w2cf3pcs、w2m3p29c、w2m3p29cs (可供第一、二波比較)

w2 高中職/專科學生 綜合分析能力(testing) v2.0：更新權數-- w2stwt1a、w2stwt1b

w2 高中職/專科老師 v2.0：更新權數-- w2stwt1



研究計畫調查資料管理

王文心

日前在網路上搜尋質性資料庫資料時，無意間看到英國「經濟與科學資料服務」(Economic and Social Data Service, 簡稱 ESDS)，針對必須或可能將調查資料交由資料庫數位保存 (digital preservation) 的學者提供「資料管理指導手冊」(Guide to Good Practice: Data Management) 供其參考，看過後覺得其內容應也有助於國內執行調查研究計畫者管理調查資料，故將一些重點整理出來，讀者可依據文中提到的項目，隨時檢視自身管理狀況，以防調查資料損毀，並確保日後的可用性。

本文提到的資料管理 (data management)，專指研究執行期間研究團隊應該注意的事項 (與資料庫長久保存資料的考量會略有差異)，包含的範圍有「哪些需要管理」以及「要如何管理」兩方面，以下是一些共通的原則，研究團隊可視本身或個別研究計畫需要發展出不同的管理策略：

- 備份檔案
- 轉換成適當的格式
- 確保備份檔案與工作檔案的正確性並做好使用管控
- 做好版本管理
- 選擇合適的儲存資料媒體與場所
- 做好資料安全管理

資料管理的前題是要有好品質的資料，研究者可以用複查、重覆輸入資料、進行不合理值及邏輯檢誤等方式來確保資料的正確性，讀者可至「學術調查研究資料庫」網站參閱「相關資源」的文章學習相關技巧。以下針對上述管理原則進一步說明。

備份檔案

常見儲存數位資料的媒體有磁片、光碟 (CD 和 DVD)、磁碟機 (隨身式、固定式) 等，這些媒體都有損毀的可能，因此在做備份計畫時，要對隨時可能發生的誤刪、誤改、儲存媒體或應用軟體出問題、電腦病毒、駭客入侵、以及人為或天然災害所造成的資料遺失等規劃好應變方案，幾方面提醒讀者特別注意：

- 備份頻率：視資料重要性、更動頻率，決定適當的備份週期
- 循環備份 (rolling back-up)：不要將新的備份檔案直接覆蓋舊有的備份檔案
- 異地儲存：理想上至少要有一份備份資料是放在不同的地點，避免災害 (例如火災、淹水) 造成備份資料連同作業檔案全部付之一炬。

- 瞭解機構備份政策：各研究室所隸屬的學校或研究機構資訊中心通常也會針對其管理的磁區進行備份工作，研究者應瞭解有哪些資料是資訊中心會定期備份的？備份的方式為何？並據此評估在這些磁區中有哪些重要的檔案需要自行額外備份。
- 備份好的檔案要進行檢查，確定能夠使用
- 選擇較可信賴的儲存媒體，例如光碟片或磁碟機。
- 更新儲存媒體：定期更換儲存舊備份資料的媒體，才能確保舊備份資料能夠讀取。
- 注意儲存媒體的環境，避免高溫、潮濕的環境，依照製造廠商的建議保存。

格式的轉換與選擇

就「學術調查研究資料庫」目前收集的台灣地區調查資料中，常見的統計軟體有 SPSS、STATA 及 SAS 三種，資料格式則有純文字檔（ASCII）、EXCEL 或上述統計軟體的系統檔。然而，不同軟體在處理資料的限制上會有稍許差異，例如選項數值說明（value labels）的部份，SPSS 和 STATA 是和數值資料結合在一個檔，SAS 則需要關聯至另一個格式檔（format file）；又如變項數的限制，STATA 9.1 標準版（intercooled）只能讀取 2,047 個變項數，專業版（SE）才能處理到 32,767 個變項，可是購買所需費用也相對的貴許多，而 SAS（9.13 service pack 3）即可處理 32,767 個變項。

不可避免地，有些資訊會在轉檔的過程中喪失，為了將轉換資料格式時可能發生的錯誤降到最低，應由熟悉資料及轉換方式的人來進行，並在轉檔之後進行檢查。主檔的備份（master copies）格式，最好不要採用非自由軟體（proprietary format）格式，如此也將會較適合日後要負責長期數位典藏之資料庫所使用的格式（見下表）；另外，應隨時注意資訊技術的發展，檢視舊有備份檔案的存檔格式是否還能被新使用的軟、硬體所讀取，進行風險評估並於必要時採取行動，避免將來無法讀取或是得付出更高的處理代價。

資料類型	計畫常見格式	資料庫長期保存格式
資料檔	<ul style="list-style-type: none"> ● EXCEL、Access ● SPSS、SAS 或 STATA 系統檔 	純文字檔+SPSS、SAS 或 STATA 欄位定義程式
問卷等文件檔	Word、HTML	PDF、純文字
數位影像	JPG、TIFF	TIFF

確保備份檔案與工作檔案的正確性並做好使用管控

數位資料容易複製、被修改或刪除，因此如何確認備份資料與主檔（master

files) 一致，並訂定一套取用管理的規定，亦是資料管理的一門重要課題，我們可以將處理過程中常會接觸的檔案分成以下幾種類型：

- 主檔 (master file)：主檔是用來儲存最完整與長期不需要更新的主要資料檔
- 異動檔 (transaction file)：異動檔是相對於主檔由某些欄位所組成的檔案，內容較少、具有時效性，只要具備所需異動的鍵值欄 (key field) 與變動項目即可
- 備份檔 (backup file)：為了避免某些有價值的檔案遭到損害，通常會將這些檔案多複製幾份，稱為備份檔。
- 臨時檔或稱工作檔 (temporary file)：資料處理過程中產生的檔案

主檔以及經過整理格式化、檢查過無誤的主檔備份 (master copies) 必須妥善保存；再者，宜依計畫執行過程的各個階段，按規劃定時備份檔案，並將之與執行計畫作業用的臨時檔案加以區隔。

維護主檔 (master files) 的幾種方式：

- 從研究團隊中指派專人負責維護原始檔案
- 依研究團隊成員工作職掌，分別給予不同的檔案讀寫權限
- 建立刪改主檔的操作步驟
- 記錄主檔更改記錄
- 保留舊的主檔，萬一新的主檔有錯誤時還可補救

做好版本管理

執行一些檢查程序，確定其中一個檔案內的資訊有所變動時，以不同格式備份的檔案、文件，以及檔案之間有互相對應的資訊都能同步更新，幾個方式提供參考：

- 記錄版本及檔案狀態，例如版本編號，檔案是屬於草稿、暫時檔或是最後定版
- 只針對一個具有適當格式的主檔進行維護，即可避免因為多個工作檔並存而產生的版本管理問題
- 記錄項目之間的關聯性：常見的狀況是，某個檔案中的資訊需要參照其他檔案的資訊，例如：編碼簿和與之對應的資料檔數值說明，或是文件檔和與其關聯的後設資料 (metadata)。
- 若儲存在不同位置，以追蹤搜尋方式找到全部有該項目的位置，同步進行更新。

選擇合適的儲存資料媒體與場所

不同的資料儲存媒體，可以儲存的資料容量從數 MB 到數十 GB 都有，配套使用的軟、硬體也會隨之不同，衍生而出的存放條件和儲存要求也會不同。最令人頭疼的是，往往儲存資料的媒體還堪用，這些軟硬體就已經被淘汰而沒辦法讀取資料。例如現在已經很少人使用的磁碟片，目前已很難找得到可以讀取 5.25 吋磁片資料的磁碟機了；奉勸手邊還有資料放在這種磁片的讀者，盡早將這些資料轉存到其他媒體中。

近年較被廣為使用的大量儲存裝置，有磁性儲存媒體和光學儲存媒體兩種，前者包括磁帶、磁碟片（軟碟）和硬碟；後者則有 CD-ROM、CD-R、CD-RW（可覆寫）和 DVD 等各式光碟片。由於製造儲存媒體所用的原料和製程技術會影響可用年限，使用時盡量選擇品質較好、信譽較佳的廠牌和供應商。查閱廠商建議的讀寫次數，使用頻率太高，拿取過程中人為不小心摔壞、刮傷，或是機器讀取資料碰觸媒體，都容易造成資料的毀損。

另外，還得注意存放環境的溫度和濕度，一般而言，溫度愈高、濕度愈大使用年限就愈短，英國標準協會（BS 4783）建議光碟片長期保存時，溫度要控制在 18-22℃，相對濕度（RH）則控制在 35-45%之間。美國國家媒體實驗室（US National Media Laboratory）另提出一份數據，在適當使用（從未使用或不常使用）以及控制好環境（避免塵污、紫外線及靠近強力磁場）前提下，媒體的壽命如下表：

	25RH/10℃	30RH/15℃	40RH/20℃	50RH/25℃	50RH/28℃
D3 磁帶	50 年	25 年	15 年	3 年	1 年
CD\DVD	75 年	40 年	20 年	10 年	2 年
CD-ROM	30 年	15 年	3 年	9 個月	3 個月

選擇儲存媒體及存放場所時要注意的重點如下：

- 保持存放及使用區域的潔淨，清除煙塵、髒污。
- 存放磁性媒體時要遠離可能產生強力磁場的區域，例如電器設備或馬達等。
- 運送磁帶時，磁帶盒或外包裝要留與外界有 5 公分的距離
- 存放在低溫、乾燥、穩定且安全的場所
- 使用前熟悉媒體的使用方式
- 使用高品質的儲存媒體及讀取裝置
- 定期保養讀取裝置並保持清潔
- 使用光碟片專用標籤貼紙及標示筆，避免標籤接著劑或墨水滲透進去傷及底下的反射層及染料層，造成光碟無法讀取。
- 盡量避免取用資料庫用來長久儲存資料的媒體，做好使用記錄，必要時進行更換。

- 製作資料庫備份資料時，建議採用二種以上的裝置和軟體/格式，儲存媒體最好也能選擇不同的廠牌。

做好資料安全管理

依「資訊技術系統 (IT systems)」及「實體 (physical) 與環境安全」兩方面，將注意事項分別列舉如下：

- **資訊技術**
 1. **網路安全**
 - 限制存取檔案權限
 - 機密資料不可存放在作為網路或電子郵件伺服器的主機中
 - 具敏感性的文件不要存在有連接網路的電腦中
 2. **升級與修正 (patches)**
 - 升級或修正與作業系統或應用軟體安全性有關時應盡速執行
 3. **病毒**
 - 所有電腦應定期更新防毒軟體及病毒碼
- **實體及環境安全**
 1. 工作人員離開時應將門鎖鎖上
 2. 放置電腦的地方應有門禁管理，儲存資料的媒體僅有特定人士可以取用
 3. 記錄被帶出電腦媒體或書面文件，登記鑰匙持有人

建議進行管理的項目

- **資料檔 (data)**

由於資料類型和產生的方式非常多樣，很難對資料的結構做出一個放諸四海皆準的定義。在社會科學方面，量化資料通常包括一或多個由數字或文字組成的長方形矩陣 (rectangular matrices)，並且有些矩陣資料可透過命令欄 (common fields/keys) 相互關聯；質性資料則通常包括訪談稿，以及聲音或影像記錄等，這些資料並不會全然皆為數位化資料。

以國內社會科學界常使用的 SPSS 為例，用 SPSS 建構的資料檔除了資料矩陣之外，還可以包含變項描述、選項數值描述、變項層次的註解等，成為深入到變項層次的後設資料 (variable-level metadata)。

- **程式檔 (code/common files)**

指能在統計分析、資料庫或試算表 (spreadsheet) 中執行命令的檔案，這些檔案會用在核對、檢查、更正、記錄、計算衍生變項、遺失資料插補、重建資料、或

是合併不同的資料檔，應一併納入資料管理的範圍中。

- **文件 (documentation)**

說明資料如何產生、數位化、及進行檢查和處理的相關文件，每一個資料檔的內容都應有細到個別變項的清楚說明。資料收集好之後進行加權、建構、插補等處理所產生的變項，也應以文件詳細記載產生的步驟以供日後查詢，若有產生這些變項的相關程式，則可併入文件中加以管理。

- **非數位化資料 (non-digital material)**

除了上述項目之外，計畫執行過程中通常還會產生一些不是用數位媒體儲存的素材，最常見的像是書面資料，例如：照片、報告書、問卷、訪談稿，以及錄音帶、錄影帶等，計畫除了要做好檔案管理外，最好盡量能將這些素材數位化以方便管理。

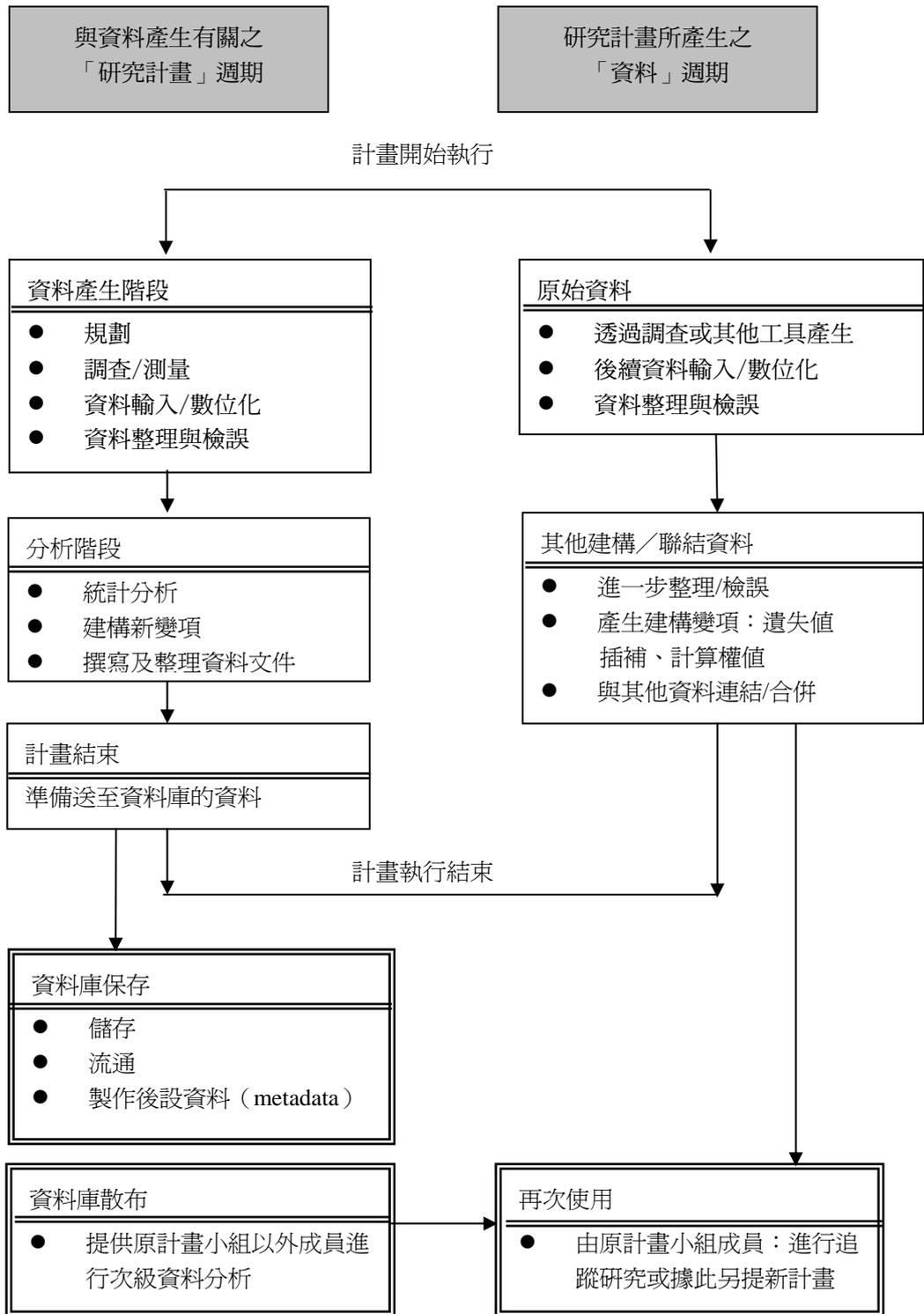
存放這些紙類物品前先做好清楚的標示，避免用釘書針、迴紋針或橡皮筋，以防鏽蝕或變質後造成污損，另外，注意存放地點的濕度、溫度、安全性、訂定借用方式等，其考量的項目與標準與數位化資料是一樣的。

資料管理與計畫執行週期

資料 (dataset) 的「壽命」通常比研究計畫久，當研究計畫因經費補助結束而停止後，研究團隊還是會持續地使用資料，也有可能另再申請研究經費，對受訪者進行追蹤訪問，或是利用前次收集的資料進行分析撰寫論文。

國外主要提供研究經費的機構，例如英國 ESRC (Economic and Social Research Council) 或美國的 NSF (National Science Foundation) 和 NIH (National Institutes of Health) 都有制定相關資料共享政策，要求經費申請者提供原始資料予公開資料庫 (public archive) 收藏，以供日後他人研究使用。在台灣，國家科學委員會也於民國 86 年時訂定同樣的政策，要求凡是以問卷進行量化資料蒐集的專題研究計畫主持人，應於計畫結束後提供調查資料檔案予中央研究院「學術調查研究資料庫」保管與釋出，希望能促進本土性調查資料庫的發展，並鼓勵學者們使用次級資料，提高調查資料的附加價值。

「資料」與產生資料的「研究計畫」週期可用下圖來表示，雙重框線的部份通常是由資料庫 (data archive) 來處理，也就是說，當計畫執行結束後，主要就由資料庫接手扮演資料保存 (preservation) 和散布 (dissemination) 的管理者角色。另外要強調的是，研究團隊準備資料提供給資料庫的整理工作最好能在計畫結束時就完成，不然極容易失去整理的動機、需要的人力也可能會因計畫結束而不足，致使資料品質達不到資料庫要求的標準。



最後，您可以試著回答下列問題，檢查看看您是否已經做好資料管理：

- 有使用標準化且一致的方法及程序來收集、檢查及驗證資料的正確性嗎？
- 是否已提供次級資料使用者關於研究方法、資料蒐集及處理方式的足夠資訊？
- 您的資料和檔案都已有完整的標示及說明，易於辨識及尋找？
- 是儲存成最新建議使用的資料格式嗎？
- 可辨識受訪者身份的隱私資料是否已經刪除？
- 書面或已數位化的備份資料是否存放在安全的場所？
- 需要備份的檔案已經完整的備份？有沒有遺漏的資訊？
- 您能清楚區分哪一個檔案是主檔嗎？
- 統計分析時可能用得上的相關資訊，例如建構變項或開放記錄等，是否也納入管理的範圍了嗎？
- 您知道誰有這些資料的智慧財產權嗎？

參考資料

1. *Guidance on Data Management* (The series of ESDS Guides). ESDS Access and Preservation, Economic and Social Data Service, UK Data Archive. Available at <http://relu.esds.ac.uk/reluaug2006.pdf>
2. *Guidance on Data Management*. Rural Economy and Land Use Programme Data Support Service. UK Data Archive. Available at <http://www.relu.ac.uk/about/data.htm>
3. *Preservation Management of Digital Materials, Ch5 Media and Formats*. Digital Preservation Coalition, UK. Available at <http://www.dpconline.org/graphics/medfor/index.html>
4. Mitchell, M. N. (2005). *Strategically using General Purpose Statistics Packages: A Look at Stata, SAS and SPSS*. (Technical Report Series, Report Number 1, Version Number 1). Statistical Consulting Group: UCLA Academic Technology Services. Available at <http://www.ats.ucla.edu/stat/technicalreports/>
5. 中央研究院調查研究專題中心（民 96），國科會專題研究計畫調查資料：資料項目與格式範例說明，台北。



台灣教育長期追蹤資料庫 (TEPS) 資料使用常見問答集

蘇婉雯

『台灣教育長期追蹤資料庫〈TEPS〉』自資料釋出以來，詢問相當踴躍。在資料內涵、研究設計、釋出規劃上的疑惑，轉給 TEPS 計畫中心後皆能獲得進一步詳盡的說明。因此本期通訊中特別整理自資料庫 TEPS 討論區中的常見問題，提供有興趣使用 TEPS 資料者參考，更竭誠歡迎您提供資料使用上的疑慮。

一、綜合問題¹

A1. 資料庫的會員，能否直接下載 TEPS 的資料？

調研中心回覆：由於申請資格與審查方式不同，TEPS 的帳號與資料庫的會員兩者是各自獨立的系統。因此資料庫的會員如需利用 TEPS 的資料，請依規定申請 TEPS 的帳號。

A2. 關於學生綜合能力分析評量，可否提供題目與學生編號，以便進一步分析學生表現與其他因素的相關性？

TEPS 計畫中心回覆：TEPS 的學生資料檔中，已提供學生表現 (IRT) 給使用者利用。至於公共使用版提供的學生綜合能力分析評量檔，僅作為 IRT score 相關研究之用，該項研究無須與學生或其他檔案進行連結，因此並未規劃釋出題目或學生編號等資訊。

A3. TEPS 第一波的綜合分析有分數學、語文、科學等題目，但為何釋出資料只有綜合與數學的 IRT 分數，而沒有語文或科學的個別 IRT 分數或答對題數？

TEPS 計畫中心回覆：TEPS 綜合分析的語文或科學等題目，設計之初就不是用來判斷學生該科目的學習成就，而是用來瞭解學生解決問題的能力〈詳見心理計量報告：分析能力測驗，第二章〉。例如：作文時字辭優美的學生，不見得在綜合分析的國文題目中拿到高分，而文采平平，但有良好解決能力的學生，仍然可以拿到高分。最主要的因素是效度與信度。如果一個變項的測量標的與研究者的研究構想不同，效度就會有問題。即使沒有效度的問題，但如果信度有問題，那麼無論分析的結果如何，都可能得到錯誤的推論。我們在釋出資料時，就因為考慮過，釋出這兩種科目的能力估計值資料，使用者可能作這樣不當的分析，而得到錯誤的推論。因此，目前 TEPS 不考慮釋出語文、科學等題目的答對題數，我們相信如此並不會減少 TEPS 的應用範圍。

數學是一個例外狀況，原因有二：第一個是，同年級學生的數學學習成就差異，並不是反映所接觸課程的不同，而是反映學生自己解決問題的能力，這正是 TEPS 的數學分析題目所要測量的。第二，數學的問題解決能力是許多

¹ 各分類中的問題順序，是依照題問次序由最近到最早排序。

實證研究特別拿來作為一般分析能力的主要指標，而我們的心理計量分析也確認了 TEPS 的數學 IRT 分數有很好的信度。所以 TEPS 額外的提供數學的 IRT 分數。

TEPS 的科任老師問卷，原本設計也不是用來瞭解學生的學習成就，除了數學的學習成就外，即使沒有其他科目的 IRT 分數，科任老師資料還是有許多方面的應用可以讓使用者發揮。

A4. 兩波可比較的 IRT 是什麼？有無使用建議？

TEPS 計畫中心回覆：所謂一、二波可比較之 IRT 〈有綜合分析能力、一般分析能力和數學三種〉，是利用在不同試卷中重複出現的題目作為基礎量尺，將沒有重複出現的其他題目，都連結在這個量尺上；再依學生在各個試卷中的表現，算出任一位學生在任一個測驗的能力估計值。這樣一來，一方面可以比較出其他沒有重複出現的試題的難易程度，另一方面尤其可以達到瞭解學生在任一階段的相對位置，並瞭解這兩階段裡，學生「能力」的起落。

一般來說，有兩種使用方法：一是將兩波可比較的 IRT 分數分別當成自變項與依變項來分析；另一種方法是計算兩變項的差異〈兩變項相減〉，即可得到學生的進步或退步的狀況。但是，在釋出資料時，還是有責任提醒使用者，有一些學生在第二波時也許不夠用心〈變項:w2illp 潛在變項模型分析下，第二波可能不用心作答的機率〉，因此他們的資料也許不能反映其實際的能力，而需要謹慎考慮是否使用這些資料。

A5. 學生跟老師檔中，有不同的加權變項〈如 w1stwt1、w1stwt1a、w1stwt1b、w1cwt1〉，請問這些變項在使用上有何差異呢？

TEPS 計畫中心回覆：基本上 w1stwt1、w1stwt1a、w1stwt1b 都是學生樣本權數，只是在不同釋出檔案，因抽取的子樣本數不同，而產生標準化過後的數值不同，故以不同的變項名稱區分；而 w1cwt1 則是班級樣本權數，只有在以班級為單位的資料檔中才有。

A6. TEPS 網站有釋出國中"導師問卷、共科老師問卷、老師問卷"之題本，但是對照資料庫後發現僅以"老師問卷"來給"導師、共科老師、老師"填答，那當初設計出三份針對不同老師的問卷目的在於？還是說，在導師問卷、共科導師問卷之目的只擷取對"學生評量"這部份來使用而已？〈資料庫獨立出學生評量的檔案〉。

TEPS 計畫中心回覆：以第一波國中為例，老師所需要填寫的問卷有三種：

- A. 老師問卷〈老師個人基本資料〉
- B. 導師問卷〈導師帶此班級的方式與感想、對此班級學生的評量〉
- C. 科任老師問卷〈國、英、數老師帶此班級的方式與感想、對此班級學生的評量〉

相對應於 TEPS 公共使用版釋出的資料，A 問卷的資料在釋出檔案國中老師共同問卷 (w1_j_t_v4_0.sav 與 w1_j_t_v4_0(stata).exe)；B、C 有關對班級的評量，因為保密性的考量，只在限制版與現場版釋出。而 B、C 關於對學生的評量是在國中學生表現評量 (w1_j_sev_v2.0.sav)。

A7. 承 A6. 是否導師需填答 2 份問卷〈導師問卷、老師問卷〉？但這些問卷有些內容又有部份重複

TEPS 計畫中心回覆：舉個例子說明：如果一位老師身兼導師與兩班的國文老師，則需填寫一份老師問卷、一份導師問卷與兩份國文老師問卷，也就是我們要蒐集這位老師的個人基本資料、導師班的班級與學生評量，還有兩個國文班的班級與學生評量。

A8. 承 A7，依導師問卷的內容，想確認是否是針對"部分問題學生"才需填答？以及另一份共科老師隨機抽 15 名的問卷，如果有受測，這些結果呈現在哪？還是說尚未釋出？因為學生對這些"受測樣本"之結果較感興趣。

TEPS 計畫中心回覆：因為 TEPS 的樣本大多是每班隨機抽 15 位學生做調查，所以導師或科任老師也就是針對這 15 位學生來評量，除非是該班抽樣學生超過 15 人，才會從中只抽 15 位來評量。

A9. 在"第二波"國中受測之導師問卷與共科老師問卷，兩份問卷僅第四部份"任教班及情形"有所差異，然而這部分在資料庫卻沒有此檔案（只呈現前三部份的受測結果）因此想詢問是尚未釋出來是其他因素？

TEPS 計畫中心回覆：TEPS 在第二波之後的調查，為了簡化問卷的種類，所以將第一波所謂的老師問卷與導/科任老師問卷合併；當然如果一位老師身兼導師與科任老師，還是兩種問卷都填寫，但 1~3 部分，只需寫一次即可。而第四部份的資料，仍然是只在限制版與現場版提供使用。

A10. 在公共使用版中，以學生問卷為例。若我 delete 部分的學生（如沒跟父母親住在一起），那麼學生的 weighting variable 是否還能使用，若否，我該怎麼做呢？

TEPS 計畫中心回覆：如果樣本數因種種原因減少了，只需將原本資料檔的權數重新標準化即可（若是使用 STATA 軟體，則不需做此動作，程式會自動幫您標準化）。

A11. 當使用公共使用版的資料與限制使用版合併時，大約只有 70% 的學生能連結，此時，該如何使用加權變項呢？

TEPS 計畫中心回覆：限制版因資料安全的考量，只有釋出公共版中約 70% 學生的老師資料沒錯；而檔案連結之後，應使用限制版檔案中的學生樣本權數將其標準化即可！

A12. 關於國中生第一、二波完訪的人數為 20,004、18,903 人，但卻只釋出 13,978、

13,247。請問，是為什麼呢？

TEPS 計畫中心回覆：TEPS 的資料目前尚無規劃將所有資料釋出，各版本均為抽樣檔，釋出比例資訊請參考『教育追蹤資料使用手冊』P.18 表 5-2 以及 P.52 5.7.4 各釋出版本資料筆數說明。

A13. 不管是國中還是高中職，家長的資料都缺乏地區別，沒有像老師的 data 有分都市、城鎮、鄉村的，是我漏看了，還是真的沒有？

TEPS 計畫中心回覆：目前釋出的家長資料中，沒有「城鄉地區分層別」的資料。不過，家長資料檔與學生資料檔可以依據 STUD_ID 此一變項合併，併檔後再依據學生資料檔中的 URBAN3（城鄉地區分層別）進行分析。

A14. 有關地區別的問題---就是地區別鄉村、城鎮、都市的操作型定義為何？是受試者主觀認定？還是依我國地方自治制度的劃分？另外，在施測問卷上有寫出來嗎？

TEPS 計畫中心回覆：地區別一共分成三類，分別是：

- 一、鄉村--〈山地鄉鎮、坡地鄉鎮、偏遠鄉鎮〉。
- 二、都市--〈工商市鎮、台北市、高雄市、省轄市〉。
- 三、城鎮--〈新興鄉鎮、綜合性鄉鎮、服務性鄉鎮〉。

其細類是依據羅啓宏「台灣鄉鎮發展類型之研究」之分層原則（共七層），再加上台北市、高雄市、省轄市等三層，共計十層分類。

A15. 想請問關於家長的出生年是民國還是西元紀元？因為用兩者算出來的結果都會有很奇怪的值。另外第二波的高中職家長問卷中，未出現家長教育程度的資料呢？如果是因為第一波已經詢問，那麼可以請問一下，我該以哪個變項為基準，合併第一波及第二波的資料呢

TEPS 計畫中心回覆：(1)TEPS 家長問卷上的原始題目為「請問您是/民國那一年出生的？」，所以問的是民國年；但家長年齡要多小或多大才視為不合理，計畫中心無法自行認定一個明確的年齡點，因此除了針對"非常特殊"的年齡〈如：負數〉做不合理值的過錄外，並不會對家長年齡做其他的處理，待由使用者自行判斷與運用。

(2)TEPS 為了能更廣泛的收集更多資訊，因此在第二波的問題題目設計上，就不重複選擇第一波已經問過的受訪者基本背景問題，使用者可利用 stud_id 此變項，來作不同波次問資料的連結，以獲得所需的資訊。

A16. 請問什麼時候會開放學校資料？

TEPS 計畫中心回覆：因為學校問卷的資料，涉及較敏感的問題，因此 TEPS 計畫中心規劃至少是在今年〈2007〉底完成所有四波的調查之後，才會進行學校問卷資料的釋出。

二、各版本問題

【公共使用版】

B1. 請問公共使用版的資料中，以學生為單位的老師資料如何與學生資料進行連結？

TEPS 計畫中心回覆：公共使用版的老師問卷資料檔，並未提供學生編號，因此無法與學生資料進行連結。在限制使用版的老師資料中，才有提供學生編號可與公共使用版的學生資料進行連結。

B2. 公共使用版所提供的綜合能力分析評量檔案，能否跟學生資料進行連結？

TEPS 計畫中心回覆：綜合能力分析評量檔案所提供的內容是綜合分析各題項的答對與答錯資訊，作為 IRT 研究者進行分析之用。同一學程不同波次的綜合分析資料則可進行連結，但無學生編號，因此無法與公共使用版中其他資料做連結。

【限制使用版】

B3. 限制使用版的老師資料與公共使用版的學生資料進行連結後，如何判定哪些學生是相同班級的？或者相同學校？

TEPS 計畫中心回覆：限制使用版的檔案中，並無提供班級編號或學校編號，因此無法進行班級或學校層級的分析。由於部分學校分科上課的情況相當普遍，試圖依據老師資料區分班級可能會產生錯誤的結果，因此不建議使用。

現場使用版提供了班級代碼或學校代碼的變項，如需進行班級或學校層級的分析，可申請現場使用版的資料。

【現場使用版】

B4. 現場使用版為何無法預約？

調研中心回覆：現場使用版僅提供網路預約使用，預約規定如下：

1. 每個月計有 1~15 日及 16~31 日兩個「使用週期」，分兩梯次開放預約登記：
 - 〈 I 〉 每月 1 日 0 時~15 日 24 時：登記當月 16~31 日使用時段。
 - 〈 II 〉 每月 16 日 0 時~31 日 24 時：登記次月 1~15 日使用時段。
 每一個使用週期內，同一申請編號以上述方式預約使用的時段，最多以 3 個為限。
2. 預約後未到的次數累計達 3 次以上之申請編號，將限制使用資格半年。無法於預定時間前來者，請務必於前一個工作天來電告知由本中心代為取消申請。

請確認申請帳號是否取得、帳號是否逾期、未到次數是否超過 3 次、當週期預約次數是否已達 3 次等情況。如有任何疑問，請來電由專人為您處理，電話：(02)27884188 分機 209。

B5. 為什麼簽了那麼多保密合約，還是不能把資料帶回去，只能現場使用？

TEPS 計畫中心回覆：在國外釋出保密資料，大致上分為兩類，一類是高保密性的資料（如人口普查資料），通常是指定地點使用（現場使用）；另一類樣本較小的保密性資料，釋出單位會不定期突擊檢查使用者是否善盡保管的責任；但 TEPS 與調研中心均無此人力，且國內法律對於使用者保密合約的規範還不夠完善與嚴謹，所以 TEPS 目前只好選擇現場使用的方式來釋出資料。即使 TEPS 的老師，一樣是需要到現場使用。

另外，依據現場使用規定，資料檔與程式均不能帶離調查研究專題中心，能帶離的檔案或書面資料僅限於『不含受訪者隱私之分析報表』（所有檔案或文件攜離前，需經由管理人員逐一檢查確認）。

因此建議現場使用版的使用者，請先下載『虛擬版』預先寫好併檔程式、分析語法等，可以節省不少於現場摸索檔案內容的時間。

B6. 現場使用版可以派助理過來嗎（申請者本人不過來）？

TEPS 計畫中心回覆：請於申請帳號時，填寫『共同資料使用者（計畫助理）』表單，每個申請人最多可填寫兩位助理，限表單上之相關人員可至調研中心使用相關資料。

參考資料來源暨相關網站：

台灣教育長期追蹤資料庫〈<http://www.teps.sinica.edu.tw/>〉

學術調查研究資料庫〈<http://srda.sinica.edu.tw>〉

會員人數

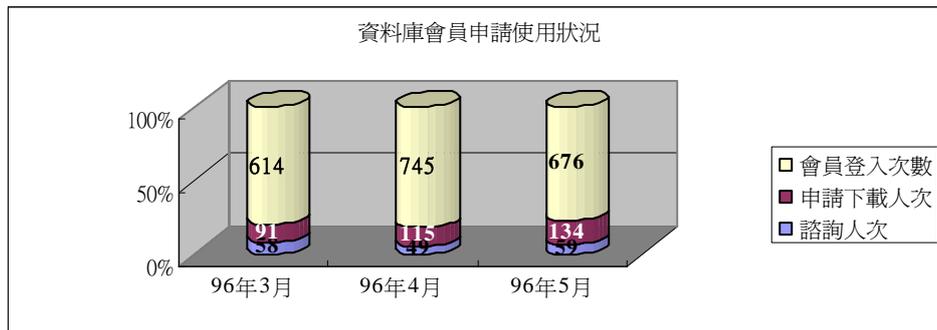
中央研究院人社中心調查研究專題中心「學術調查研究資料庫」目前的會員人數共 1484 人，其中包括一般會員 743 人，以及臨時會員 741（有效數字至九十六年五月底）。

一般會員包括：國內外公私立研究機構研究人員 83 人；國內外公私立大專院校教師 620 人；政府機構相關研究人員 4 人；捐贈資料單位代表 8 人；調研中心工作人員 28 人。

臨時會員則包括了：國內外公私立大專院校大學部、碩、博士班在學學生 702 人；國內外公私立研究機構臨時性研究人員 22 人；國內外公私立大專院校臨時性研究人員 16 人；政府機構臨時性研究人員 1 人。

資料下載、諮詢人次及會員登入次數統計

本資料庫會員於本資料庫網頁登入之後，可以隨時下載資料並瀏覽計畫相關文件；若有特殊需求或該釋出資料有特殊釋出申請的限制，則須與工作人員聯絡並透過申請的方式取得；使用者詢問有關資料釋出相關問題、檔案使用方式、程式語法等，經由專人解答之後，即為資料庫處理的「諮詢」項目。九十六年三至五月底會員登入次數、申請下載人次及諮詢人次，如下圖：



大型主題計畫下載人次統計（有效數字自九十六年三月至五月底）

- ◆ 台灣地區社會意向調查——10 人次
- ◆ 台灣地區社會變遷基本調查——189 人次
- ◆ 台灣地區家庭與生育力調查系列——11 人次
- ◆ 台灣選舉調查資料——49 人次
- ◆ 國民營養健康狀況變遷調查——27 人次
- ◆ 華人家庭動態資料庫——167 人次



會員申請程序說明

一、會員資格

(一) 一般會員

1. 國內外公私立研究機構研究人員；
2. 國內外公私立大專院校教師；
3. 政府機構相關研究人員；
4. 捐贈或授權資料予「學術調查研究資料庫」之個人或單位代表。

(二) 臨時會員

1. 國內外公私立大專院校大學部、碩、博士班在學學生；
2. 國內外公私立研究機構臨時性研究人員；
3. 國內外公私立大專院校臨時性研究人員；
4. 政府機構臨時性研究人員。

二、會員福利

1. 線上資料下載；
2. 線上瀏覽報告書、問卷、次數分配；
3. 定期寄贈「學術調查研究資料庫通訊」；
4. 電子報服務；
5. 調查資料整理與檢誤諮詢服務；
6. 各項網頁功能。

三、加入方式

(一) 上網登錄

請至本資料庫網頁「會員專區」讀取「會員約定條款」並填寫申請表，送出資料後將螢幕上顯示的個人資料表列印出來，在劃線處簽名以示同意該條款中的各項規定，並請將該份書面資料以郵遞方式擲回本中心存查。

- 提供文件：在職（在學）文件影本。
- 審核時間：收到文件後一至三個工作天。

(二) 親自申請

親至本資料庫辦公室填寫申請表格，交付相關文件並簽署同意會員約定條款。

- 提供文件：在職（在學）文件影本。
- 審核時間：視申請人身份而定。符合者可以立即生效。

四、資格認證

- (一) 一般會員：會員有效期間為期兩年，以成功加入會員日起計算。本資料庫將每兩年主動校對會員資料，若會員未於校正期間內確認資料異動情況，則將先暫停其會員權限。
-

- (二) 臨時會員：會員有效期間為期一年，以成功加入會員日起計算。如仍符合臨時會員身份資格者，應每年於期限前主動提出證明文件，得延長使用期限一年。

五、會員義務

資料庫在努力提供更好的服務的同時，亦需要了解會員們的意見。因此有一些需履行之義務項目，請會員配合實行。

- (一) 資料使用：敬請遵守「會員約定條款」的各項規定。
- (二) 著作回覆：會員在下載資料使用後，發表相關篇章論著等，敬請提供書目資料，嘉惠他人作為研究方向的輔助參考之用。若能提供作品全文亦非常歡迎。
- (三) 意見調查
1. 當次：會員於下載資料三週後，資料庫會主動針對會員下載記錄詢問該次下載之資料用途，以及資料的使用滿意度；
 2. 年度：針對會員一年來的使用情形，本資料庫將進行例行性的年度整體滿意度調查，以作為資料庫發展及改善的參考。



資料捐贈說明

本資料庫亦歡迎學者及各界捐贈資料！為使這些得來不易的珍貴資料發揮最大效用，本資料庫工作人員會謹慎處理每筆捐贈資料，並負起妥善保管之責。我們蒐集的資料主要是以調查方法進行的量化資料，公開釋出前，我們會先瞭解該筆資料之智慧財產權歸屬，並取得所有權人/機構同意授權本資料庫公開學術使用。

本資料庫對資料捐贈者提供以下服務項目：

1. 定期／不定期寄贈「學術調查研究資料庫」出版品；
2. 資料整理與檢誤相關諮詢服務；
3. 提供其捐贈資料之申請情況；
4. 提供資料使用者利用其捐贈資料發表之衍生著作。

若您願意捐贈資料給「學術調查研究資料庫」蒐藏並對外開放，或有任何疑問，歡迎洽詢「中央研究院調查研究專題中心」資料組，我們將竭誠為您服務。

電話：(02)27884188 分機 201，資料組組長王文心小姐。

E-mail: srda@gate.sinica.edu.tw



徵稿啓事

「『學術調查研究資料庫』通訊」為中央研究院人文社會科學研究中心調查研究專題中心出版以服務學術界之刊物，每年三月、六月、九月及十二月以季刊形式發行。本中心為增加本通訊內容之多樣性及促進與讀者的互動交流，歡迎學界人士能夠提供以下相關議題之文章。

- 一、 二手資料之利用與分析
- 二、 調查資料之整理與檢誤
- 三、 國內外資料庫使用之經驗
- 四、 資料編碼或編碼簿之製作
- 五、 統計軟體之運用

寫作重點請著重於經驗交流與心得建議。請以 e-mail 的方式傳寄給我們，字數以 1000-1500 字為限，並以單篇完結之形式寫作。來稿由編輯委員會審閱決定刊登與否，一經採用將稿費酬謝，每千字 750 元。對接受刊登稿件，本刊編輯群有權刪改並置放於「『學術調查研究資料庫』通訊」電子版網站，如不同意者，請於稿件上聲明。文稿內容不代表本刊意見，如涉及著作權或其他文責問題，由作者自負法律責任。文章以真實姓名發表，投稿請附上您的聯絡方法（通訊地址與電話）。

『學術調查研究資料庫』通訊

Survey Research Data Archive Newsletter

出版單位／中央研究院人文社會科學研究中心 調查研究專題中心

地 址／(115)台北市南港區研究院路二段 128 號

電 話／(02)27884188

傳 真／(02)27881740

網 址／<http://srda.sinca.edu.tw>

電子郵件／srda@gate.sinca.edu.tw

發行人／傅祖壇

編輯顧問／于若蓉、杜素豪、林季平、楊孟麗、廖培珊

主 編／王文心

文 編／李孟諳、邱亦秀、黃瓊瑤、楊景閔、楊惠婷、廖軒琦、蘇婉雯

美 編／邱亦秀

印 刷／阜橋國際有限公司

地 址／(115)台北市忠孝東路六段 36 號 1 樓

中華民國九十一年四月創刊

中華民國九十六年六月第廿一期

GPN : 2009101543