

# 「學術調查研究資料庫」九十年度工作成果報告

王文心

## 一、前言

中央研究院調查研究工作室於民國 83 年 11 月成立「學術調查研究資料庫」，由資料管理組負責相關業務之規劃、執行與發展，並開始針對台灣地區以面訪、電話訪問、郵寄問卷等調查方式所蒐集到的學術調查資料，進行有系統的蒐集、整理、保管與釋出工作。

目前「學術調查研究資料庫」的調查資料來源，以本室協辦調查業務、政府機關授權提供學術界使用，以及學者熱心捐贈等為主，累計蒐集 600 多筆的資料；完成整理與檢誤，並已屆計畫主持人同意公開期限而對外釋出的資料將近 350 筆。資料庫開放學術界人士申請使用以來，本組提供服務的申請及諮詢人次，各有 2406 及 699 人次（統計至 90 年底）。

除上述例行工作之外，本組也參與委託本室協辦調查研究計畫之資料整理與檢誤的工作，以及統籌採購與管理本院人文社科各所同仁在研究上需求的數據型電子資料，並負責舉辦本室重要學術活動--「調查研究方法研習班」。

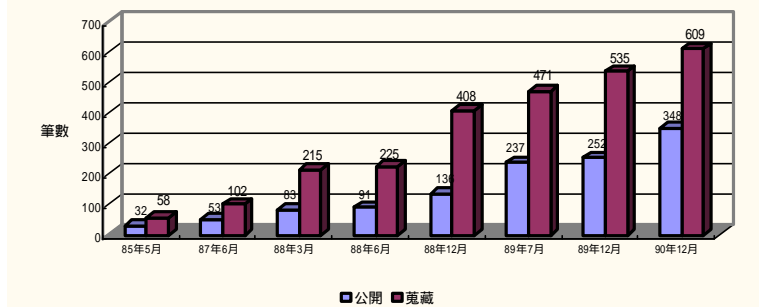
「學術調查研究資料庫」共有全職人員七名，另外雇用兼任助理與工讀生各一名，協處理資料檢誤與行政業務。在全職人員專長方面，一位具圖書資料管理背景，負責資料的編目、管理與處理申請、諮詢案件，其他六位的背景包括公衛、社會、心理、社工等，其皆受過量化資料處理訓練，並熟悉各種統計軟體的應用，為本組負責資料整理與檢誤、協辦計畫的主力。

在本文中，我們將著重在介紹九十年度資料蒐集、整理與檢誤及釋出申請的情形，以期藉由此文，讓學術界人士對「學術調查研究資料庫」的運作能有更深一層的認識。

## 二、調查資料蒐集狀況

在九十年度中資料庫共蒐集到 55 筆調查資料，資料來源主要是國科會委託管理之國科會專題研究計畫調查資料，計有 52 筆（94.5%），其他 3 筆（5.5%）為本室協辦計畫的調查資料。累計至本年底，資料庫共蒐藏 609 筆調查資料，其中整理檢誤完畢，並已屆計畫主持人同意公開日期而對外釋出者，有 348 筆（圖一）。

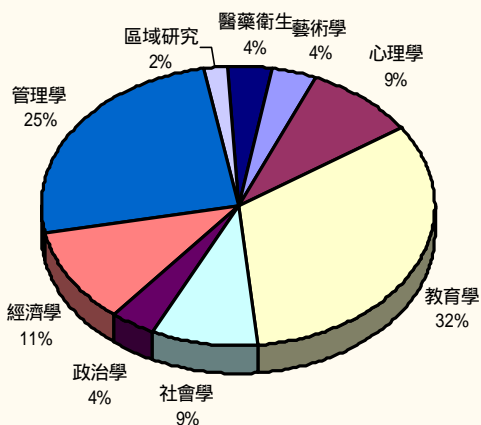
圖一 歷年調查資料蒐集與公開情形



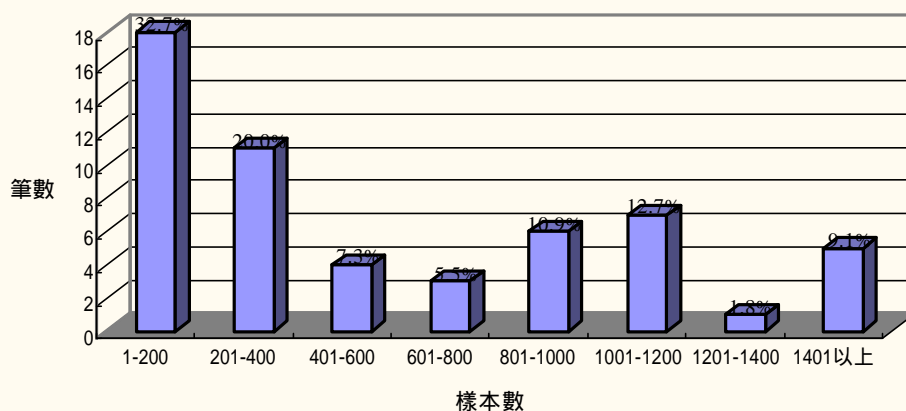
九十年度新蒐集到的資料在學門分布上，以教育學、管理學、經濟學及社會學為主，依序占九十年度收入調查資料的 32%、25%、11%及 9%（見圖二）。調查方式以郵寄方式占大部份，約一半左右（50.9%），其次為面對面訪問，占約四分之一（25.5%），其他則為電話訪問（9.1%）、自填問卷（18.2%）等。母體地區以全國性調查為主（53%），其次為區域性調查（20%），

值得一提的是有 2%是在外國地區進行調查。在調查週期方面，則是以橫斷性調查為主（80.0%），跨年追蹤調查、一年內連續追蹤調查以及未說明調查週期者，依序為 16.4%、1.8%及 1.8%。最後看到成功樣本數時（圖三），若以 1000 個樣本做為分界，成功樣本數在 1000 以下者所占的比例（76.4%），高於成功樣本數在 1000 以上者（23.6%）。

圖二 九十年度收入之調查資料學門分布(N=55)



圖三 九十年度收入調查資料樣本數分布情形(N=55)



依資料庫蒐藏調查資料的標準，以及因應電子化資料庫的需求，一項完整的調查資料應包含的項目如表一所示。在九十年度中，資料庫蒐集到的資料在完整度上比往年進步。就整體而

言，資料庫蒐集到的九十年度調查資料中，以取得過錄編碼簿及報告書電子檔的難度較高，各有 18.2%與 20.0%只能蒐集到書面資料或甚至連書面資料皆無。

表一 學術調查研究資料庫對外釋出之調查資料標準

資料項目	書面資料	電子檔	說明
資料讀我檔	必要	必要	檢誤時依原計畫提供之資訊，分別製做中文與英文讀我檔案。
空白問卷	必要	必要	
過錄編碼簿	必要	必要	必須包含題號、欄位、變項名稱、變項說明、選項數值說明。
資料檔	--	必要	提供系統檔者，可不須提供欄位定義程式。
欄位定義程式	必要	必要	必須包含欄位讀取、變項名稱、變項說明、選項數值說明。
報告書	必要	必要	

在過錄編碼簿所需要包含的資訊中，以題號、欄位、選項數值說明等三項比較不易取得，原始資料未有這三項說明的比例依序為 13.0%、6.5%及 4.3%，但在變項名稱及變項說明方面，提供的比率則高達 100%。

在資料檔與欄位定義程式方面，本資料庫在初始蒐集調查資料時，便有 52 筆（94.5%）的資料提供資料檔，而有提供欄位定義程式的有 24 筆，占 43.6%。沒有提供欄位定義程式的情況有二，一為將其數據資料存成統計軟體內定的系統檔，該類資料可直接讀入使用；另一為以 MS EXCEL 建檔，分析時僅以 EXCEL 進行計算，並未使用統計軟體。然而，不論是使用系統檔或欄

位定義程式，資料庫都會希望其中能包括變項名稱、變項說明、選項數值說明的資訊，以利分析結果報表的閱讀。本年度新進資料在變項名稱、變項說明、選項數值說明這些項目的完整度依序為 95.9%、75.5%及 67.3%。大多數學者（90%）採用 SPSS 統計軟體為其分析資料的工具。

對於上述無法蒐集到完整調查資料項目或內容資訊的資料，我們會再與計畫主持人聯絡，請其協助提供完整資料或資訊，如此能便利我們在整理對外釋出資料的統一格式時利用，再者亦可增加釋出資料的完整度。而若是無法取得計畫主持人的完整資料或資訊時，則由本室人員依據蒐集到的資料逕行整

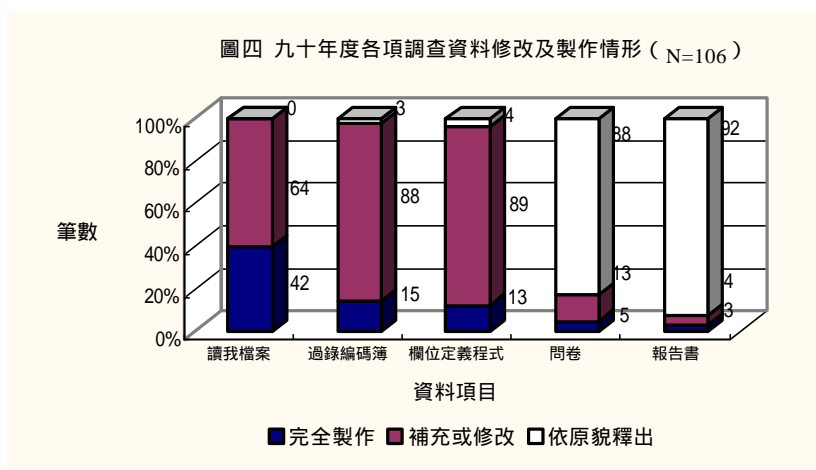
理，並將整理好的調查資料寄送計畫主持人做最後確認之後，本室才對外公開資料提供有興趣者申請。

### 三、資料整理與檢誤

資料庫會依據資料預定對外公開的時間以及資料的完整度安排整理與檢誤的作業，在九十年度時，共對 106 筆調查資料進行整理與檢誤。

為了統一資料對外釋出的格式，增加使用二手資料的便利性，我們依本

資料庫資料釋出標準格式重製或修改計畫主持人原先所提供的部份資料內容。本資料庫以計畫主持人所提供的資料或資訊分別製作成本資料釋出格式的中、英文讀我檔案；空白問卷、報告書若無電子檔時，則由本資料庫重新繕打；而在過錄編碼簿及欄位定義程式（系統檔）方面，為符合資料庫的釋出格式，而需要重製或修改的比例則在九成以上（圖四）。



在細部內容的處理方面，過錄編碼簿及欄位定義程式除了各有 3 筆及 4 筆資料完整，不需要任何修改外，其餘需要修改的資料中，尤以欄位定義程式中缺少選項數值說明的情形最為普遍，需由工作人員補充的比例高達 83.3%。過錄編碼簿中需要工作人員修改題號、欄位、變項名稱、變項說明、選項數值說明的比例依序為 53.4%、51.5%、56.3%、69.9%、68.0%；欄位定義程式中需要修改欄位定義、變項名稱、變項說明、更

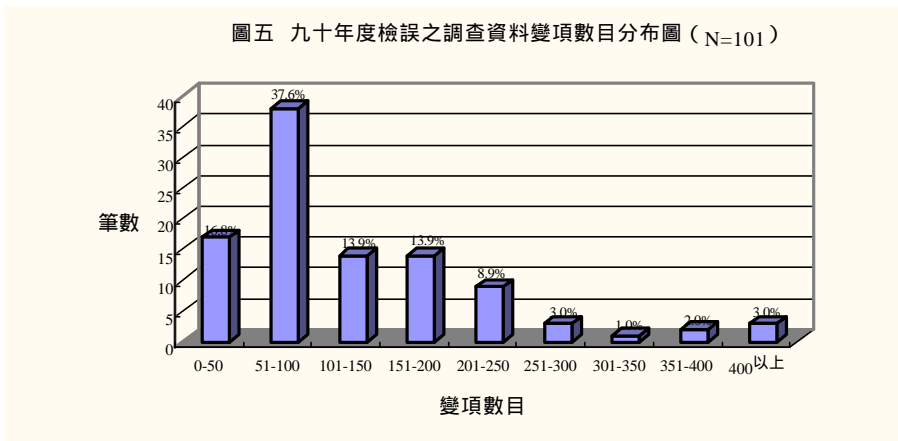
正程式語法的比例，則依序為 27.5%、25.5%、70.6%、37.3%。

在資料檔處理方面，因為檢查出在資料檔中含有不合理值 (illegal value)，在計畫主持人授權之下由本資料庫修改資料檔的比例最高，占 20.8%；其次為轉換資料格式 (例如：將 Excel 檔轉換成 ASCII 或是 SPSS 系統檔、STATA 檔轉 SPSS 系統檔等)，占 15.1%；其中有 7 筆 (6.6%) 資料因為計畫主持人未將可辨識受訪者資料

的變項刪除，而為保障受訪者權利，我們將這些變項刪除後再對外釋出。

從變項數多寡上，大致可看出問卷的長短，而變項數也與處理資料所需的時間有關，變項愈多者，整理資料所需要花費的時間也就愈久，扣除到九十

底年還未完成最後檢誤之 5 筆資料後，本年度所檢誤的調查資料變項數目分布情形如圖五所示，101 筆資料中有 58 筆（57.4%）出現不合理值，有不合理值的變項從 1 個到 258 個不等（表二）。



因為報告書內容的詳盡程度不一，檢誤的資料中有 67 筆因報告書中沒有次數分配結果，因而未能加以核對外，在另外 39 筆載有次數分配結果的資料中，有超過一半 (n=21, 53.6%) 的研究計畫資料檔次數分配結果與報

告書記載的一致 (表三)，不一致的變項數在 10 個以內的占 30.8%，而不一致的變項數目有 11、17、22、及 29 個變項的計畫各有一筆，並有 2 筆計畫報告書記載的樣本數與資料檔的不符。

表二 檢誤之調查資料中出現不合理值的數目

變項數目	筆數	百分比
0	43	42.6
1~5	39	38.6
6~10	5	5.0
11~20	3	3.0
21~50	3	3.0
≥51	8	7.9
合計	101	100

表三 檢誤結果與報告書次數分配不一致之變項數目

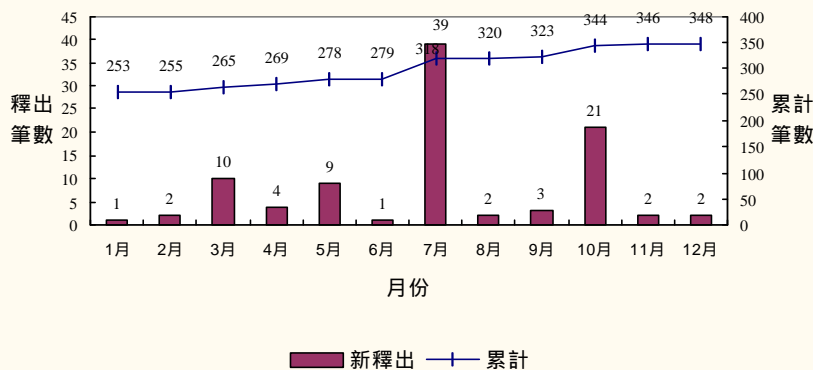
變項數目	筆數	百分比
0	21	53.8
1~5	9	23.1
6~10	3	7.7
≥11	4	10.3
樣本數不一致	2	5.1
合計	39	100

#### 四、釋出資料情況

資料整理與檢誤完成後，經過計畫主持人確認釋出資料內容，並已過計畫主持人指定之公開期限後，本組便在中研院週報以及學術調查研究資料庫網頁上公布釋出消息。國科會計畫資料的公開期限，在八十八年度以前是以計畫執行結束後最晚二年公開，八十九年度第一期之後，是由計畫主持人指定之，本室協辦計畫的調查資料，則是最晚於計畫執行結束後五年公開。

在釋出方式上，本年度亦有兩項重要變革。首先，我們將國科會專題研究計畫的問卷轉成 PDF 檔後，置於網頁上，方便使用者瞭解調查的內容，縮短找到符合需要資料的時間；另外，在聯合報民意調查方面，一改以往採年度為單位對外釋出的方式，改以單筆資料為釋出單位，並將歷年來（民國 76 年-86 年）共計 1029 次調查之調查地區、樣本數、主題表列於網站上，提供有興趣申請使用者做為參考。

圖六 九十年各月份釋出資料量與累計量分布圖



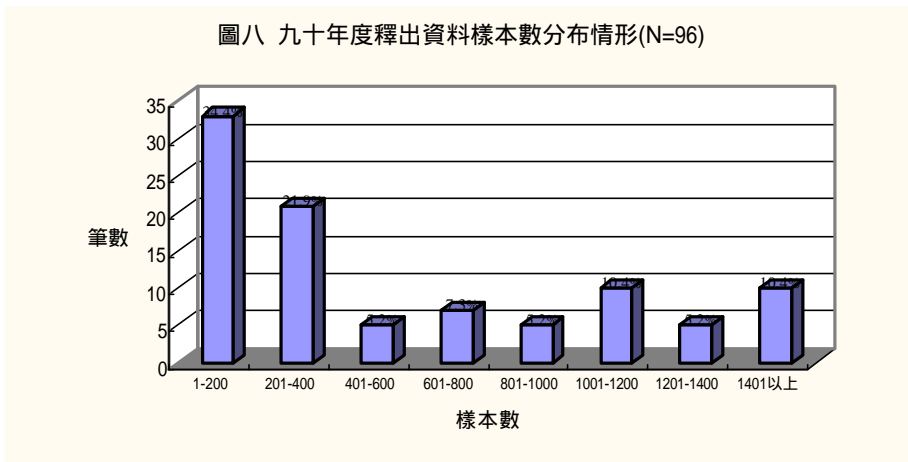
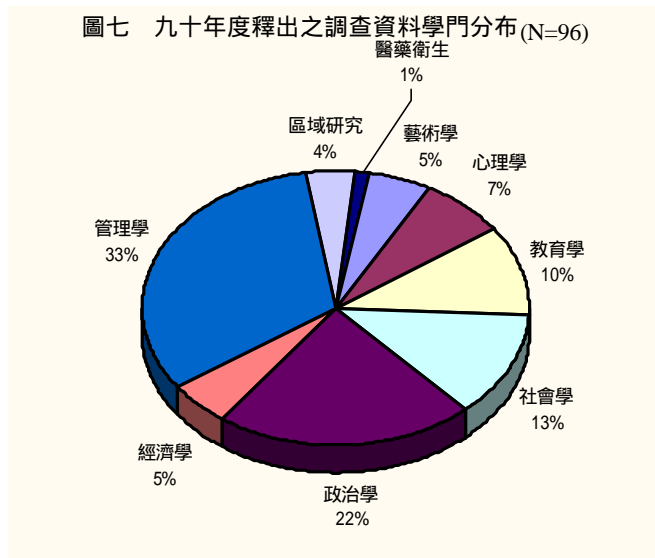
九十年度內共新釋出 96 筆資料，累計釋出量在九十年底時達到 348 筆。因國科會計畫執行結束日期通常為 7 月底，故在 7 月時會有較大量的資料釋出（圖六）。

釋出資料中最多的是管理學（33%）與政治學（22%）的調查資料（圖七），並以面訪（49.0%）、郵寄問卷（37.5%）、自填問卷（17.7%）進行

者占多數。在其他有關調查方式的相關資訊方面，調查母體地區中以全國性的調查占 40%，其次為區域性調查（25%）、單一縣市調查（18%）；抽樣方式中，機率抽樣占 69.8%、採非機率抽樣並註明原因者占 16.7%、而未說明採非機率抽樣原因或未說明抽樣方式者，依序占 6.3%與 7.3%，最後在長期性調查方面，跨年追蹤調查的資料占

5.2%，橫斷性調查資料占 89.6%，無法得知者占 5.2%。至於樣本數的分布方

面，有一半左右（56.3%）的資料樣本數在 400 以內（圖八）。



值得一提的是，在本年度釋出資料中有兩項大型資料，一為台大醫學院生化科黃伯超教授與中央研究院生物醫學科學研究所潘文涵研究員主持的「國民營養健康狀況變遷調查」，該調查以經過訓練之專職人員，於民國 82 年 7 月到 85 年 6 月間，對全國四歲以上的

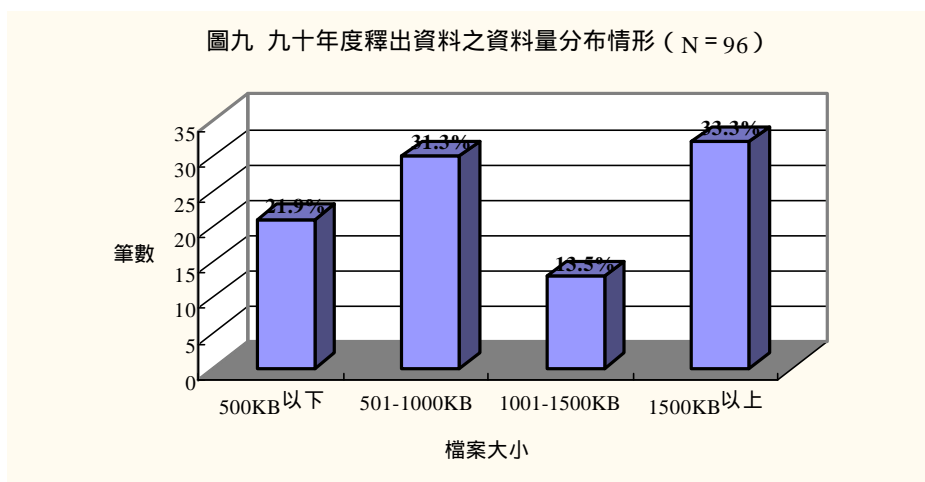
民眾，分年齡性別層，進行飲食、營養、健康狀況的整體調查。

另一大型資料為台大政治系歷年選舉研究計畫，這是一系列從民國六十六年到民國八十六年針對各項總統、立法委員、省市長選舉議題進行的調查，

共計有 19 筆，主要採面訪的方式進行調查。

為了便利使用，除系統檔外，我們都會附上欄位定義程式，以便使用者拿到資料後，便可直接進行分析。檢誤時我們沿用計畫主持人原來使用的統計軟體進行程式的修改，釋出資料所使

用的統計軟體中，仍以 SPSS 最多，其次為 SAS，其他則有 STATA、MINITAB 等。各筆資料包括讀我檔案、問卷、資料檔、過錄編碼簿、報告書、欄位定義程式等及其他資料的檔案大小分布，則如圖九所示。



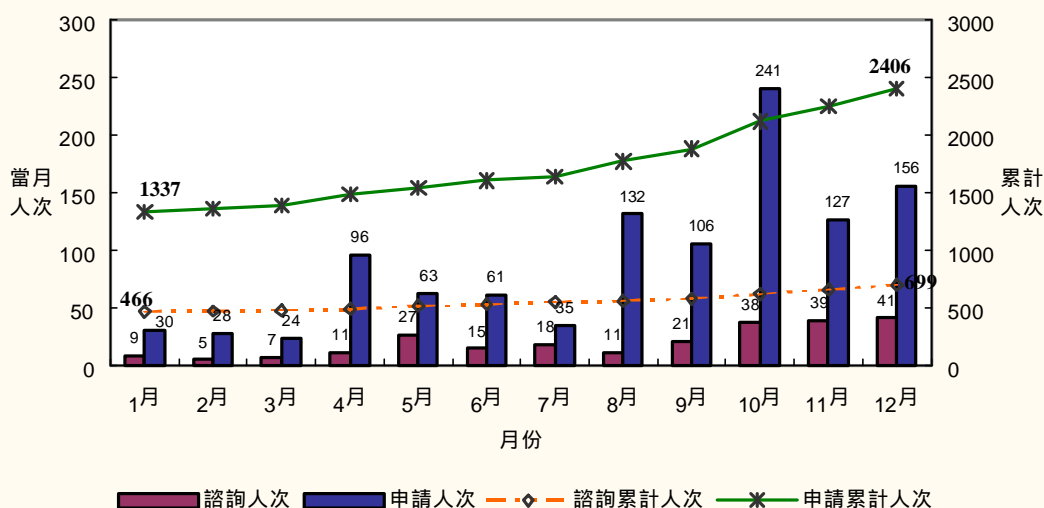
## 五、申請人次

「學術調查研究資料庫」自 84 年 11 月起對外開放，提供二手資料予學術界，凡是國內外公私立研究構內的研究人員、大專院校老師、碩博士班學生、在老師指導下的大學部學生、及其他經資料提供者同意者，皆可申請使用「學

術調查研究資料庫」之調查資料。在九十年度中，共服務的資料申請和諮詢人次依序為 1099 位與 242 位，各月份申請與諮詢人次，以及累積申請與諮詢人次如圖十所示。從資料庫開放使用至九十年底，申請與諮詢人次已累積到 2406 人次及 699 人次。



圖十 九十年各月份申請與諮詢人次



申請者的身份以研究生（47.59%）和大專院校教師（24.02%）為主（表四），申請次數最多的資料是歷年的台

灣地區社會變遷調查資料（表五），各月份的分布請見圖十一及圖十二。

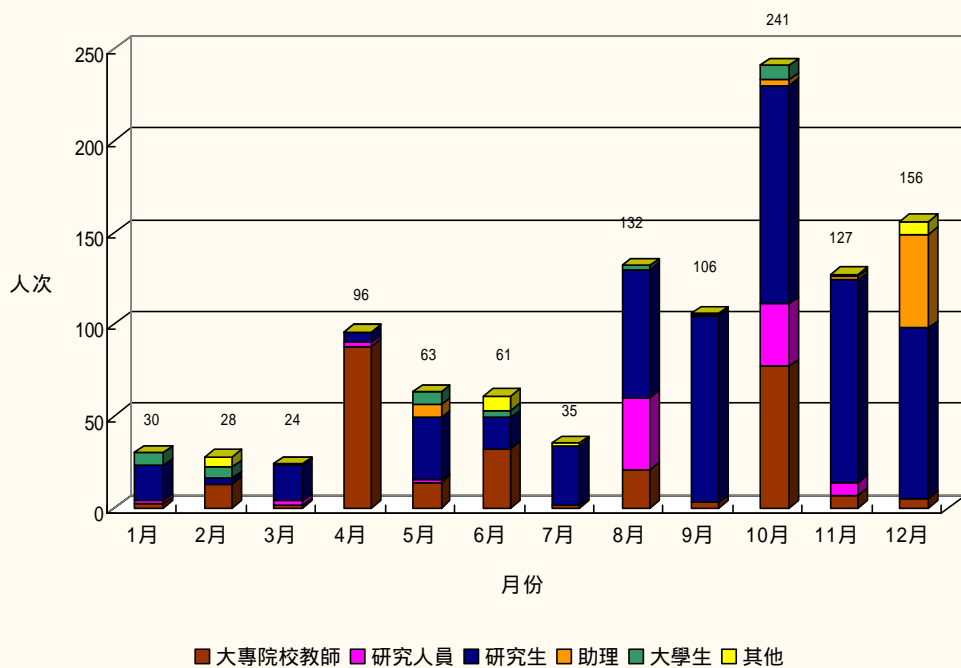
表四 九十年度資料申請者身份別分布情形

申請身份別	人次	百分比
大專院校教師	264	24.0
研究機構研究人員	88	8.0
碩、博士研究生	624	56.8
研究助理	63	5.7
大學生	37	3.4
其他	23	2.1
合計	1099	100

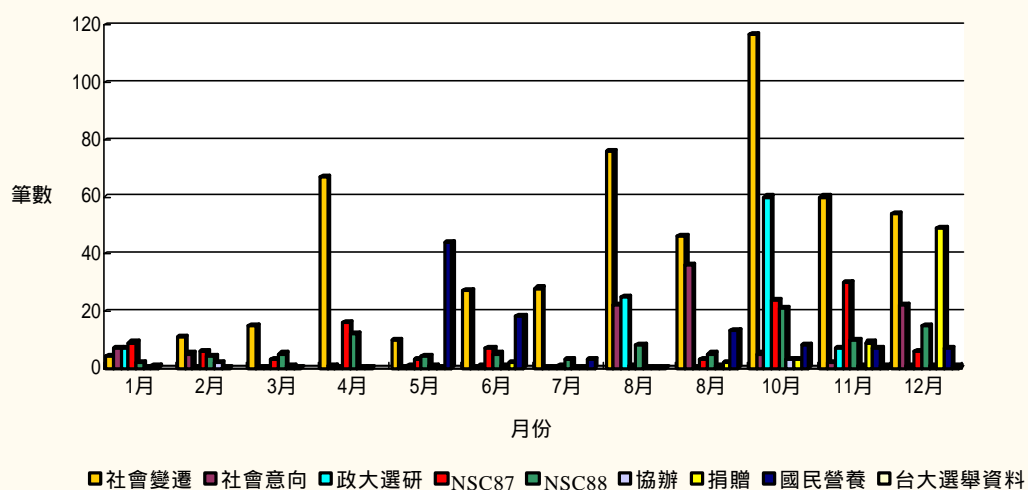
表五 九十年度申請使用之資料類型分布情形

申請資料類別	人次	百分比
社會變遷	515	46.9
社會意向	100	9.1
政大選研	102	9.3
NSC87	109	9.9
NSC88	94	8.6
協辦	11	1.0
捐贈	66	6.0
國民營養	100	9.1
台大選舉資料	2	0.2
合計	1099	100

圖十一 九十年度各月份申請人數及身份別分布圖 (N = 1099)



圖十二 九十年度各月份申請資料類別 (N = 1099)



其他重要工作成果包括：

1. 協辦計畫：

協助辦理院內同仁委託進行之調查研究計畫的資料整理與檢誤、分析資料、及撰寫執行報告書等。九十年年度中，共計協辦 10 項次。

2. 舉辦調查研究方法研習班：

在九十年九月份，分別邀請政大社會系關秉寅教授講授「調查研究方法導論」、台灣大學政治系洪永泰教授講授「抽樣理論與方法」、以及本室田芳華助研究員講授「調查問卷資料處理與檢誤」。

3. 外購數據型電子資料庫：

負責採購與管理本院人文社科各所同仁在研究上需求的數據型電子資料，九十年度優先購置的數據型電子資料有三：

- (1) 行政院主計處人力資源調查暨附帶專案調查資料；
- (2) 美國標準普爾公司 (Standard & Poor) 發行之 COMPUSTAT 資料庫；
- (3) 美國芝加哥大學商學院發行之 CRSP 資料庫。

## 六、結語

由以上的報告可以看出，九十年年度中「學術調查研究資料庫」是以調查資料整理與檢誤為工作主軸，持續地蒐集與釋出資料，以奠定「學術調查研究資料庫」成為國內調查資料數量最多、學科領域最廣之專業資料庫的基礎。除此之外，亦加強服務本院同仁的研究需求，購入國內外數據型的電子資料庫供研究人員使用；並舉辦「調查研究方法研習班」，提供院內同仁及國內大專院校師生吸取調查研究新知的機會。

在九十年度中，我們已經著手規劃加強資料釋出服務的部份，冀望利用網路無時間、空間限制的特性，將更多有用的訊息透過資訊科技來傳達。而目前已完成的工作項目有調查資料問卷、調查資料整理與檢誤講義、與學術調查研究資料庫使用手冊的上網作業；爾後，我們更將朝全文檢索與提供線上分析的方向發展。我們將秉持對學術服務的熱忱，繼續蒐集與整理更多調查資料供學界使用、開發使用者便捷取用調查資料的平台，並和學界分享我們在資料處理技術上累積的實務經驗。希望藉由這些努力，能夠不斷地自我充實與提昇，以期能與國際接軌，並對國內學術界做更大的貢獻。