

教育資料與圖書館學

*Journal of Educational Media & Library Sciences*

<http://joemls.tku.edu.tw>

---

Vol. 55 , no. 1 (2018) : 39-69

台灣社會科學量化資料

再用之研究：2001-2015年

The Reuse of Quantitative Data in  
Social Sciences in Taiwan: 2001-2015

林奇秀\* Chi-Shiou Lin \*

Associate Professor

E-mail : [chishioulin@ntu.edu.tw](mailto:chishioulin@ntu.edu.tw)

賴璟毅 Ching-Yi Lai

Graduate Student

**[English Abstract & Summary see link](#)**

**[at the end of this article](#)**





# 台灣社會科學量化資料 再用之研究：2001-2015年

林奇秀<sup>a\*</sup> 賴璟毅<sup>b</sup>

研究論文

## 摘要

本文分析2001-2015年間，社會科學領域利用既有量化資料從事具原創性研究之狀況，取TSSCI 2015年收錄之經濟、政治、社會、教育與心理共五學門57種期刊，辨識出1,484篇再用論文，佔整體實徵論文17.38%，經濟與政治學門比例最高，心理則非常稀少。全部被引資料筆數共2,990筆，亦多由經濟與政治論文使用，且兩大學門論文同時使用多筆資料為常態；社會與教育論文則以使用一筆資料居多。在資料來源與資料類型上，政府機關資料及業務數據為被用大宗，但不同領域在資料來源與類型上有明顯差異；個別研究與一次性資料再用程度都很低。本文建議社會科學的資料度用服務應著力於蒐集並媒合公私部門的大規模、長期性數據資料，並建立跨政府與民間的服務機制，當可裨益資料的充分再用。

**關鍵詞：**資料再用，資料引用，量化資料，社會科學

## 一、前言

資料(data)是能夠描述事物、思想、情況、處境，與其他因素的實物、數字、文字，或符號(National Research Council, 1999)，在形式上，資料可包含數字、統計、表格、圖像、影片、地理空間測量數據等(Data Information Specialist Committee - UK, 2010, as cited in Kruse & Thestrup, 2014)。資料是科學研究的基礎，對強調實徵探究的社會科學而言，無論量化或質化研究，均須仰賴資料分析來描述現象，解釋成因，進而提出研究論述或形成理論(Sechrest & Sidani, 1995)。

<sup>a</sup> 國立臺灣大學圖書資訊學系副教授

<sup>b</sup> 國立臺灣大學圖書資訊學系研究生

\* 本文主要作者兼通訊作者：chishioulin@ntu.edu.tw

2017/09/04投稿；2017/12/15修訂；2018/02/07接受



近年來資料分享 (data sharing) 呼聲漸高，鼓勵學者釋出一手資料，將之典藏於公開典藏庫，供他人再利用 (Borgman, 2012)。科學研究資料的分享呼籲在 1980 年代興起，對某些特定領域如天文與生命科學而言，科學知識的突破亟需大量數據，故不同研究團隊必須群策群力，共享珍貴研究資源 (Konkiel, 2013; MacMillan, 2014)。時入千禧年後，E-Science 與數位基礎建設 (cyberinfrastructure) 的推展，不僅改變科研學術生態，也讓資料皮用 (data curation) 成為新顯學。資料皮用指的是在研究資料產生、使用、移轉、保存、銷毀這一系列的生命週期中，透過適當的管理程序與技術對資料進行妥善的維護，並透過數位典藏庫來永久保存 (陳雪華, 2012; Data Curation Center, n.d.)，皮用目的是在確保研究資料能被發現、檢索、加值，與再利用 (Henderson & Knott, 2015)。

基於上述趨勢，近年來關於資料皮用與研究資料管理的文獻不少，但是對於學者運用既有資料來從事衍生分析的情況，實徵研究不多。再者，既有的資料再用研究多關注科學領域，對社會科學領域的探討相對有限。然而，社會科學學者的資料再用，其實是行之有年的研究行為。歐美國家自 1960 年代起，即建立多個社會科學資料典藏機構，如美國密西根大學的知名跨校資料典藏聯盟 Interuniversity Consortium for Political and Social Research (ICPSR) 即成立於當時。顯見社會科學的資料再用比科學領域的分享倡議發生得更早，資料典藏與共享體制的悠久歷史，也凸顯了資料再用在社會科學的普及性與重要性。

社會科學與自然應用科學在資料再用上，也存在著重要差異。首先，社會科學學者使用他人蒐集的個別研究資料來進行再分析的比例，其實不高，在社會科學領域中，大部分的再用資料是政府或科研機構釋出的大規模調查或業務統計數據 (林奇秀、賴璟毅, 2017; Fear, 2013)。再者，社會科學學者所使用的資料類型也不限量化數據，還包含文本、手稿、影音紀錄與實物等質性資料，但由於這些非數據資料在採集、分析、與詮釋上，與量化數據有極大差異，且非數據資料的典藏、公開與分享條件等，與數據資料也大不相同。目前，這類質性資料是否應該分享、如何皮用、再用的分析效力為何等，都仍具相當爭議 (Broom, Cheshire, & Emmison, 2009; Jeng, 2017; Kim, 2013; Van den Berg, 2005)。

目前少數可得的資料再用相關研究，多以自然科學為分析對象，分析重點大致在辨識再用論文與被引資料 (cited data) 的特徵。例如：再用論文的年代與數量成長狀況 (Piowar & Vision, 2013; Xia & Liu, 2013; Zhang, Vogeley, & Chen, 2011)、再用論文與被引資料的學科領域分布 (Fear, 2013; Piowar, 2008; Piowar, Carlson, & Vision, 2011; Robinson-García, Jiménez-Contreras, & Torres-Salinas, 2016; Zhang et al., 2011)、再用論文類型 (Fear, 2013)、再用論文的資

料使用量 (Piwowar & Vision, 2013; Zhang, 2011)、再用論文品質 (Piwowar & Vision, 2013; Xia & Liu, 2013)、再用論文作者國家別 (資料再用較頻繁的國家; Xia & Liu, 2013)、再用論文與被引資料的年代關係 (Piwowar & Vision, 2013)、再用資料的來源別 (Zhang, 2011) 等。

然而，上述研究都面臨到一項研究方法的侷限，亦即如何辨識資料再用論文。研究者除非逐篇閱覽論文，否則難以輕易辨識一篇論文所用資料，到底是該次研究中自行蒐集產生的，抑或援引自外部。現有的資料再用研究，多是利用各學科領域知名資料典藏庫或資料引文索引 (如 Web of Science 的 Data Citation Index)，以其資料辨識碼關鍵字串先檢出論文內文或參考書目含有該典藏庫資料者，再透過人工或自動的方式，篩除生產該筆資料的原始論文，留下合乎「再用」定義的論文來進行分析。透過這類辨識獲得的樣本固然合乎條件，但也可能會遺漏其他合乎再用定義的論文，例如作者並未適當引用資料、或被引資料根本未收藏在那些典藏庫中。換言之，當樣本侷限在特定典藏庫或資料引文索引，我們就無法完整得知各領域的資料再用程度與再用行為的學科差異。同時，對非英語系國家如台灣，學者再用的資料也不見得取自國際性資料典藏庫或引文索引，因此上述方法，也不適合用於探索台灣學界的狀況。

但相對於卷帙浩繁的國際性英文期刊，台灣的學術期刊總體規模小，反而可以採取逐篇瀏覽、人工辨識的方法來完整檢出再用論文，進而比較學科的再用情況。本研究故以 TSSCI 2015 年期刊收錄清單為基礎，選取社會、經濟、政治、教育，與心理共五大學門，分析 2001 至 2015 共 15 年間的資料再用狀況。具體而言，本文企圖回答的問題，包含：

- (一) 再用論文佔前述領域論文的比為何？成長趨勢為何？
- (二) 再用論文使用資料筆數的狀況為何？成長趨勢為何？
- (三) 被引資料的來源為何 (即資料的蒐集者或彙編者)？有無學科差異？
- (四) 被引資料的資料類型為何？有無學科差異？

如前所述，由於質性資料型態繁複且再用狀況與量化資料差距甚大，因此，本研究僅聚焦於量化資料再用。本研究所指量化資料，係指社會科學學者為進行實徵分析而自行蒐集或援用外部來源所得的數據型資料。被學者再用的量化資料，形式上可包含從外部取得的完整、獨立存在的資料集 (dataset)，也可以是學者從數據資料庫、統計刊物，或其他文件逐項檢出、自行彙整形成的資料集。換言之，只要資料集內的原始數據不是學者透過自行調查或實驗而產生的，其分析行為就視為資料再用。社會科學研究常用的量化資料類型，包含公私部門從事各類調查所得的原始數據，以及反映各類業務活動的業務統計 (業務統計也包含主事單位對各類自然與社會現象持續監測所得的數據，例如氣象、空氣品質、交通流量、人口統計等)。此外，雖然數量不多，也仍

有社會科學學者運用前人個別研究所得數據來進行再分析(林奇秀、賴璟毅, 2017; Fear, 2013)。

本文所指的「資料再用」, 必須是對既有的量化資料進行具原創性的衍生分析, 而非單純在驗證前人已發表的結果是否真確。資料再用必定包含原始數據的統計運算, 其目的是在驗證新的研究假設或產生新的研究發現與推論。如果學者取用數據, 僅為強化或佐證論述, 不涉及資料的重新運算與詮釋(如僅用統計數據來鋪陳研究背景、強化論點, 與其他統計數據比較等), 就僅視為一般的資訊引用, 而不視為資料再用。

本文所指的「被引資料」, 即再用論文中援引自外部的量化資料。一篇再用論文, 可使用單筆或多筆取自外部的資料集。值得注意的是, 由於各領域的寫作習慣不同, 學術引文規範對於資料引用也缺乏明確規範, 因此, 學者所援用的資料不見得都會以引註方式標註在參考書目, 因此, 本研究所指的被引資料, 泛指所有在論文本曾描述的被用資料, 不僅限於列名在參考書目者。

為行文方便, 本文後續將以「數據庫」一詞來專指資料典藏庫(data repository), 藉以與台灣中文習稱的「資料庫」(database, 泛指包含各種類型文本、資訊條目, 或數位媒材的儲存庫)作有效區別。必須注意的是, 中國大陸用語中的「數據庫」, 概念等同台灣習稱的資料庫, 與本文所指的數據庫概念並不相同。

## 二、文獻回顧

資料再用是近年才受矚目的議題, 既有實徵研究相當有限, 研究目標多以描述若干特定科學領域之再用特徵、再用行為, 及各種阻礙再用的問題為主, 且以仰賴大規模數據之科學領域為主, 如基因與天文研究, 社會科學資料再用的研究很少, 且僅少數初探性質化研究著眼於學科脈絡與資料再用的關連, 如Chao(2012)以科學研究的「時間性」(time scale)與節奏(rhythms)分析地球科學中兩種不同類型資料(實物標本及電腦運算數據)的短期與長期再用狀況。

從樣本來源來看, 既有的實徵研究多以領域內的數據庫出發, 去尋求有使用數據的再用論文, 繼而再分析數據再用與再用論文之關係與特徵; 另有一篇是以資料引文數據庫作為樣本基礎, 以下分述之。

### (一) 自然與應用科學的資料再用

#### 1. 基因領域

基因領域的再用研究有三篇。Piwowar(2008)的研究問題是基因研究資料的再用情形, 她以2007年PLoS中200篇含有「microarray」(即DNA微陣列研究)關鍵字的論文作為樣本, 以Gene Expression Omnibus(GEO)數據庫資料及

人工辨識的方式來區辨其中哪些為初用論文（亦即生產研究資料的原始論文）及再用論文。接著，再比較再用論文與初用論文使用自行產製資料的勝算比(odds ratio)，也就是一件事出現的機率跟另一件事的機率比值，計算方式是將兩者的勝算相除。結果發現：菌類研究有最高的勝算比(2.4)，而在老鼠、細菌、病毒、植物等領域，勝算比均小於0.5。換言之，在基因研究上，菌類的DNA微陣列研究論文較常再用既有資料。

Piwowar與Vision(2013)則探討有公開分享資料的初用論文能否獲得較高引用量，同時也觀察再用論文的使用狀況。在再用論文的檢索部分，他們先以主題關鍵字「gene expression microarray」(基因表現微陣列)從PubMed Central、HighWire Press及Google Scholar找到10,557篇論文，續以GEO以及ArrayExpress兩個數據庫的辨識碼關鍵字串進行比對，找出其中2,617筆有公開資料的初用論文，其後，再透過這些論文原始資料的確切識別碼去搜尋，找出曾經引註這些資料的論文，並人工刪除與初用論文作者姓名相符者，共找出4,543筆第三方再用論文(亦即再用論文與初用論文的作者全不重複)。其重要研究發現包括：(1)第三方再用論文的年代分布介於2000至2010年之間，且於2004年之後快速增長。(2)再用論文年代與資料年代差距約在3至6年之間。(3)在資料使用量上，2003至2004年的第三方再用論文多數僅使用1至2個資料集，至2005年時，最多不使用超過6個資料集，然而，到2010年時，已經有25%的論文至少使用3個以上資料集，且最高超過50個資料集。

Xia與Liu(2013)關切的是再用論文的年代分布、作者來源、發表在哪些期刊、論文品質狀況等。他們以2011年以前有使用GEO數據資料、且被收錄於PubMed的第三方再用論文為樣本，共1,211篇。其主要發現包含：(1)第三方再用論文年代分布於2003至2010年之間，年代最早者為2003年共5篇論文，之後逐年增加，至2010年該年單年已有數百篇。(2)近600筆第三方再用論文的作者來自美國，遠高於第二名的英國(近500筆)，而第三至五名的國家分別為中國、德國與日本，開發中國家幾乎都沒有再用論文。(3)在論文品質方面，這些再用論文分布在286份期刊中，但48%集中於10份生醫期刊，以兩年影響係數(2-year impact factor, IF)來觀察，全部期刊的平均IF為5左右，前十名的期刊IF均高於25，顯示第三方再用論文具有一定的品質。

## 2. 天文領域

天文學的資料再用研究有兩篇，均以Sloan Digital Sky Survey(SDSS)數據庫的使用為樣本基礎，但此處的資料再用狀況與基因研究不太一樣，GEO數據庫內的資料大多為個別研究所貢獻出的資料，而SDSS本身是一個大型的、組織性的天文觀測數據收集計畫，數據則公開提供給天文學界分析使用。

Zhang等(2011)分析1,453篇使用SDSS數據庫資料的論文，發現：(1)出版

於1994至2007年間的再用論文，數量有逐年增加的趨勢，如2005年有264篇論文，至2006年則大幅增長為423篇。(2)1994至2001年間，論文平均作者數量逐年增加，至高為17.77人，但於2002年起逐年減少，於2007年6月僅剩6.49人。他們認為研究團隊規模變化的原因，是因早期SDSS的主導議題都需要較大研究團隊，而近期學界關切的議題不像過往那麼龐大，因此作者人數就變少了。(3)在論文主題方面，他們分析論文題目與摘要包含的名詞關鍵詞(noun phrases)數量，藉以了解論文主題繁複程度變化，結果發現，1994至2007年之間，每篇論文名詞關鍵詞平均介於10至11個之間，惟1998年少於10個，而2007年則為12.82個。他們並援引生物多樣性公式計算論文主題差異，結果發現論文主題差異程度有逐年增加的趨勢，換言之，使用SDSS資料的論文，主題是愈來愈多元的。

Zhang (2011)在天文學文獻資料庫NASA Astrophysics Data System (ADS)中，以“SDSS”與“Sloan AND Survey”為關鍵字進行檢索，找出2,632筆可能有用到SDSS數據的論文，並從中隨機抽取200篇進行內容分析，其中169篇(84.5%)確實有使用SDSS資料。分析結果顯示：(1)在使用資料數方面，169篇論文總共使用了278筆資料，有98篇僅使用1筆資料，有48篇使用2筆資料，只有23篇使用3筆以上資料。(2)在資料來源方面，研究者不僅使用SDSS資料，全部辨識出47個不同資料來源，略可分為：144筆(51.9%)公開數據庫資料(含SDSS)、35筆(12.5%)僅流通於天文社群之資料、59筆(21.2%)學者自行蒐集的初次資料、38筆(13.7%)作者從先前研究累積之自有資料。

### 3. 其他科學領域

Piwowar等(2011)從三個不同學科數據庫的使用狀況來探討學科之間的資料再用差異，三個數據庫分別為基因領域的GEO、地球與環境科學的Pangaea、及種系發生學(phylogenetics)的TreeBASE。從這三個數據庫，各隨機抽取100筆2005年資料，尋找曾引用這些資料的再用論文。由於資料尚無標準引註方式，導致再用論文的辨識有相當難度，因此，他們用了兩個方法來尋找再用論文：首先，他們使用資料的DOI號碼，在Google Scholar中檢出含有資料DOI的論文，並人工確認論文確實有資料再用；其次，他們用這300筆資料的初用論文題目，在Web of Science (WoS)中檢索其後續引文，唯300篇論文的後續引文數量眾多，且不見得都是引用其原始資料，故他們先從全部引文中抽出150篇樣本，以人工辨識其中確實為資料再用論文者，再觀察三個數據庫在這些再用論文中的比例，進而以此比例來推估三個數據庫衍生出的全部資料再用論文篇數。根據上述兩個方法，他們發現：在資料使用量上，三個資料庫被使用的程度相差懸殊，如以DOI查詢結果來看，在2005至2010年間，GEO被使用了114次，Pangaea僅有7次，TreeBASE僅4次；但如根據引文來推

估，GEO估計被使用了436次，Pangaea增加到581次，而TreeBASE僅28次，顯示種系發生學的資料再用遠低於前兩個學科。

## (二) 社會科學的資料再用

Fear (2013) 的研究是在觀察社會科學領域中，學者再用個別研究產出資料的情形，不含政府或機構釋出的大規模調查或業務數據。ICPSR有一文獻資料庫稱作Data-Related Publications，蒐集曾引用ICPSR資料的文獻，在研究期間該資料庫收有2,323篇再用論文，而Fear以2000至2002年間的273筆個別研究資料去比對這些再用論文，辨識出有再用這些資料的論文。分析發現再用論文的特徵包含：1. 在論文類型方面，有1,560篇(67.3%)期刊論文，374篇(16.13%)研究報告(Report)，157篇(6.78%)書籍。2. 有1,325篇(57.04%)論文是資料的自我再用，僅577篇(24.84%)為第三方再用，其餘271篇(11.68%)則非實徵研究，故非再用。3. 在577篇第三方再用論文中，有449篇為期刊論文，以Scopus分類架構進行分類後，發現排名前五名的論文主題分別為醫學類(116筆，25.84%)、醫學與心理類(74筆，16.48%)、社會科學類(59筆，13.14%)、醫學與社會科學類(48筆，10.69%)、心理學與社會科學類(38筆，8.46%)。4. 經多元迴歸分析發現，資料作者的聲望(以作者的h-index計算)、資料處理狀態(資料是否先經過整理)、資料作者團隊規模、資料主題是否跨領域、資料規模大小(變數多寡與樣本大小)，以及相關文獻數(資料作者使用該資料的出版數量；資料被二次分析的文獻數量)等，皆對資料的再用次數有所影響。

## (三) 基於資料引文的跨學科比較

Robinson-García等(2016)以Data Citation Index (DCI)中的引用數據為基礎，分析各學科的引用狀況。DCI是WoS所含資料庫之一，內容涵蓋三種型態資料記錄：資料集(data sets)、資料論文(data studies，亦即有使用前述data sets的論文，含本文所指的初用與再用論文)，以及數據庫紀錄。2013年間，DCI收錄2,468,736筆(94.1%)資料集紀錄、154,674筆(5.9%)論文紀錄，及90筆數據庫紀錄，而該研究發現：1. DCI資料紀錄的主題分布為：科學領域2,118,855筆(80.76%)、社會科學462,862筆(17.64%)、藝術與人文51,444筆(1.96%)、工程與科技領域有1,789筆(0.07%)。值得注意的是，雖然科學領域紀錄量極多，但工程科技領域卻極少，甚至比藝術與人文類低。2. 在被引狀況上，全部2,623,528筆紀錄條目中，有88.1%未被引用，總被引次數為413,689次。整體而言，每個資料集平均被引次數為0.12，資料論文為0.69，數據庫則為36.28。3. 以學科分布來看，科學領域條目被引319,458次(79.03%)，社會科學被引72,855次(18.02%)，藝術與人文被引20,460次(5.06%)，工程與科

技被引916次(0.23%)。4.在資料集294,091被引次數中，結晶學佔所有被引量47.4%，而結晶、生化、基因學共佔被引次數86%以上。5.在資料論文122,333被引次數中，社會學佔30.8%，其次是人口學佔17%，經濟學佔14.3%。相對於資料集的引用比例，社會科學在資料論文的被引佔比較高。

### 三、研究方法

本研究以TSSCI所收錄之社會科學期刊為基礎，取2001至2015年共15年間發表的量化資料再用論文為樣本，以內容分析法來了解社會科學整體與分領域論文之資料再用狀況。TSSCI於2015年的收錄清單共收有105筆期刊，分布在九個學門(人類、社會、教育、心理、法律、政治、經濟、管理、區域研究與地理)、綜合類，及新興跨領域等11個類別中(科技部人文社會科學研究中心，2015)。本研究選擇社會、經濟、心理、政治以及教育等五個學門進行分析。

科技部的學門分類與一般認知的社會科學學科領域分類不完全一致，例如傳播學劃分於社會學門之下、圖資與體育學門劃在教育學門之下等，造成學門內領域殊異程度非常高。為深入觀察，本研究進一步觀察刊名、刊物的官方出版資訊，以及出刊內容等，進一步將某些學門內的期刊進行學科分類，以符合學科領域認識。具體來說，社會學門再分為社會學、社會工作(簡稱社工)、傳播學三個學科；政治學門再分成政治學與公共行政與國際關係(簡稱「公行國關」)兩個學科；教育學門分成教育學、圖書資訊學，與體育學三個學科；經濟學門分成經濟學與農業經濟(簡稱「農經」)兩個學科；心理學門下則僅維持心理學一個學科。

此外，原列於「綜合類」下的四份期刊，因與前述學門及學科主題關係非常密切，本刊故納入分析，包含人口學刊歸入社會學，教育與心理研究、測驗學刊、中華心理衛生學刊歸入心理學。經歸類後，本研究共分析57份期刊(參見附錄1)。

上述部分期刊的發刊年代晚於2001年，另有部分期刊在這15年間曾歷經刊名變遷，由於刊名變遷可能代表編輯政策或內容性質的更迭，因此，期刊如有改名者，均僅取現刊卷期作分析，更名前之舊刊論文均不納入。在清查論文時，本研究先去除非研究論文(如編者言、書評、投書評論、會議實錄等)，續而區辨研究論文中的實徵研究論文與非實徵論文(如理論辯證、現象剖析、理論模式或公式建構但未採用實徵資料印證者)，再以人工瀏覽所有實徵論文的摘要與內文特定區塊(如前言、研究方法、圖表註釋等處)，確認其所用量化數據是否援自外部或由該論文自行蒐集產出，並對再用論文逐一編號，續將每筆再用論文與被引資料特徵登錄於Microsoft Excel檔案中，以利後續統計分析，登錄項目含：

### 1. 每筆再用論文登錄資訊

- (1) 論文篇目資訊(含期刊卷期與出版年)。
- (2) 論文所屬學門與學科(依期刊所屬之學門及學科來判別)。
- (3) 該篇論文所引用之外部資料筆數。

### 2. 每筆被引資料登錄資訊

- (1) 資料辨識資訊項目：資料集題名、資料來源(如作者、彙編者、提供者等)、識別碼或定位資訊(如DOI、URL或資料所在位置之文字描述)、描述資訊(關於資料集的其他內容描述如樣本規模、變數項目、調查主題等)。
- (2) 資料來源：本研究將資料來源區分為政府機關、學術機構、民間單位，及個別研究(含個人學者或個別研究團隊)等四種。
- (3) 資料類型：本研究將資料類型區分為「業務資料」(指公私機構或科研單位在長期業務執行或持續觀測中所積累與成長的數據資料，例如政府機關的業務統計、環境監測數據、人口統計等)、「系列調查資料」(公私機構或政府單位基於特定研究目的而定期執行的調查資料，例如社會意向調查、選舉民意調查等調查統計)，與「一次性資料」(由個別學者、研究團隊，或公私機構基於特定研究目的，在單次研究中所蒐集的資料，且可辨識出該筆資料為一次性、無持續增補之意圖)。

## 四、研究發現

### (一) 再用論文數量比例與成長趨勢

2001-2015年間，再用論文的數量狀況詳如表1。前述57份期刊共出版有1,484篇再用論文，佔全部研究論文的11.99%，佔實徵論文的17.38%。以學門狀況來看，再用論文數量由高至低依序為經濟、政治、社會、教育，與心理學門，經濟學門佔了全部再用論文的四成(40.23%)，政治學門也佔全部再用的1/4以上(26.62%)。此外，經濟學門的實徵論文高達八成(79.6%)為再用論文；排名第二的政治學門，再用論文佔實徵論文雖不足一半，但仍達四成之高(40.79%)，顯示這兩個學門不僅是台灣社會科學領域中的資料再用大宗，學門內資料再用的現象也非常普遍，對經濟學門而言，甚至是主流的資料使用方法。

值得注意的是排名第四的教育學門，其再用論文佔實徵論文比例雖然僅比心理學門高一些(5.47%)，遠低於前三個學門，但這是因為教育學門的全體論文數量為全體學門之最。事實上，其再用論文數量實高出心理學門甚多，數量直逼社會學門，也佔全體再用論文數的15.23%。因此，如就社會科學整體來看，教育學門的資料再用是仍具相當規模的。

表1 再用論文之數量、比例，與學門／學科分布 (2001-2015)

學門／學科	研究 論文數	實徵 論文數	再用 論文數 (佔整體再 用%)	佔研 究論 文%	佔實 徵論 文%	學科 數量 排名	學科 比例 排名*
<b>整體</b>	<b>12,381</b>	<b>8,541</b>	<b>1,484</b> (100.00)	<b>11.99</b>	<b>17.38</b>		
<b>經濟學門</b>	<b>1,125</b>	<b>750</b>	<b>597</b> (40.23)	<b>53.07</b>	<b>79.60</b>		
經濟學	764	469	395	51.70	84.22	1	1
農業經濟	361	281	202	55.96	71.89	2	2
<b>政治學門</b>	<b>2,265</b>	<b>836</b>	<b>341</b> (26.62)	<b>15.06</b>	<b>40.79</b>		
政治學	1,216	442	187	15.38	42.31	4	4
公共行政與 國際關係	1,049	394	154	14.68	39.09	6	5
<b>社會學門</b>	<b>1,690</b>	<b>1,045</b>	<b>243</b> (16.37)	<b>14.38</b>	<b>23.25</b>		
社會學	602	381	195	32.39	51.18	3	3
社會工作	292	208	30	10.27	14.42	9	6
傳播學	796	456	18	2.26	3.95	10	10
<b>教育學門</b>	<b>5,304</b>	<b>4,131</b>	<b>226</b> (15.23)	<b>4.26</b>	<b>5.47</b>		
教育學	3,541	2,535	186	5.25	7.34	5	7
體育學	1,306	1,219	34	2.60	2.79	8	9
圖書資訊學	457	377	6	1.31	1.59	11	11
<b>心理學門</b>	<b>1,997</b>	<b>1,779</b>	<b>77</b> (5.19)	<b>3.86</b>	<b>4.33</b>		
心理學	1,997	1,779	77	3.86	4.33	7	8

註：\*「學科比例排名」係依學科領域內再用論文佔實徵論文之比例排序

心理學門的資料再用就非常罕見，其再用論文數量不足百篇，且佔學門實徵論文僅4.33%，比例非常低。更甚者，細觀心理學門所含期刊，有四份實與教育高度相關（中華輔導與諮商學報、教育心理學報、教育與心理研究、及測驗學刊），這四份期刊共出版了53篇再用論文，佔心理學77篇的68.83%。換言之，非屬教育相關的心理學再用論文，數量真的很低。

就學科層次來看，無論從論文數量或佔實徵論文比例來看，依序均以經濟、農經、社會、政治等四個學科為首，公行國關在數量與比例上亦居前。相對來說，社工的再用論文比例雖名列第六，但論文數量卻少，不足百篇；而教育學再用論文比例雖遠低於社工，數量排名卻名列第五。後文在分析學科層次的再用論文與被引資料特徵時，將取再用論文數量超過百篇的六大學科進行分析，故含教育，但不含社工。

圖1與圖2則顯示15年區間內，整體樣本及學門／學科領域中，再用論文佔每年實徵論文百分比的逐年分布。無論在整體樣本、學門，或六大學科內，資料再用都未見明顯的成長趨勢。尤其就整體樣本來看，再用論文每年佔比相對穩定，均維持在15%至22%之間，平均值為17%，歷年來最高值與最低值距

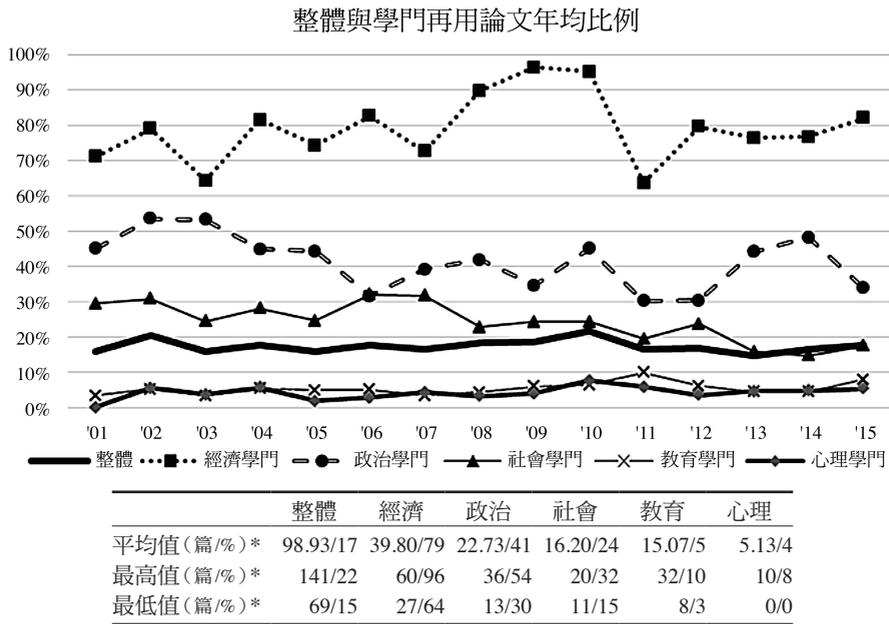


圖1 各學門再用論文比例年均分布曲線圖

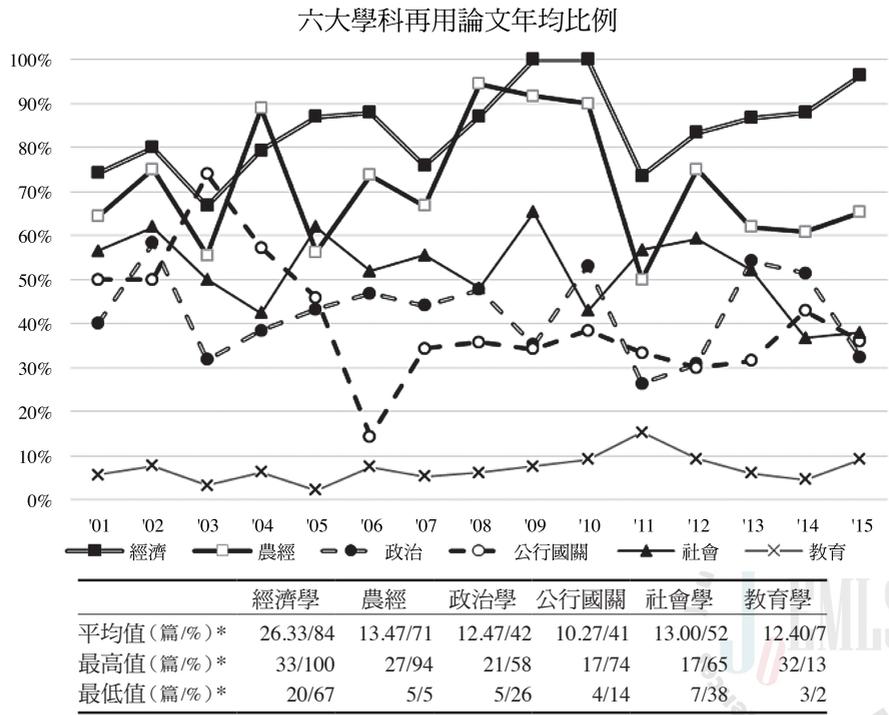


圖2 六大學科再用論文比例年均分布曲線圖

平均值約2%-5%。至於在五個學門與六大學科的年分布曲線上，經濟學門與政治學門起伏略大；在六大學科中，除教育學相對平穩，其他五個學科的起伏均大，但除經濟學有微幅上揚，其他學科並未見有明顯的上升或下降趨勢。換言之，本文所觀察的區間內，台灣社會科學領域中似無系統性驅力在強化或弱化資料再用。

## (二) 再用論文的資料使用量

2001-2015年間，再用論文的資料使用量狀況如表2。1,484篇再用論文共使用了2,990筆資料集，平均每篇使用2.01筆，標準差為2.02。在學門層次，以每篇論文平均使用資料多寡來看，依序是經濟、政治、社會、心理，與教育學門。經濟學門全部使用的資料筆數相當龐大，佔整體一半(50.30%)，且是五個學門中唯一在資料使用筆數比例上超過再用論文數量比例的，顯見其資料再用不但普及，且每篇論文使用資料筆數量也超過其他學門。政治學門使用資料數量次於經濟，約佔整體1/4(23.21%)。整體來看，經濟與政治學門相加，即佔全部被引資料約3/4，且這兩者的平均資料使用筆數均高於整體平均值。然而，政治學門的標準差大於平均值，且是五個學門中唯一者，表示該學門論文

表2 再用論文使用資料筆數概況(2001-2015)

學門/學科	再用 論文數 (佔整體 再用%)	使用 資料數 (佔整體 資料數%)	平均值	標準差	學門 排名 (依平均值)	學科 排名 (依資料數)	學科 排名 (依平均值)
<b>整體</b>	<b>1,484</b> (100.00)	<b>2,990</b> (100.00)	<b>2.01</b>	<b>2.02</b>			
<b>經濟學門</b>	<b>597</b> (40.23)	<b>1,504</b> (50.30)	<b>2.52</b>	<b>2.38</b>	<b>1</b>		
經濟學	395	1,001	2.53	2.21		1	1
農業經濟	202	503	2.49	2.68		2	2
<b>政治學門</b>	<b>341</b> (26.62)	<b>694</b> (23.21)	<b>2.04</b>	<b>2.11</b>	<b>2</b>		
政治學	187	389	2.08	2.31		3	3
公共行政 國際關係	154	305	1.98	1.85		4	5
<b>社會學門</b>	<b>243</b> (16.37)	<b>385</b> (12.88)	<b>1.58</b>	<b>1.39</b>	<b>3</b>		
社會學	195	291	1.49	1.16		5	7
社會工作	30	57	1.90	2.07		8	6
傳播學	18	37	2.06	2.07		10	4
<b>教育學門</b>	<b>226</b> (15.23)	<b>297</b> (9.93)	<b>1.31</b>	<b>1.12</b>	<b>5</b>		
教育學	186	243	1.31	1.20		6	10
體育學	34	45	1.32	0.68		9	11
圖書資訊學	6	9	1.50	0.84		11	7
<b>心理學門</b>	<b>77</b> (5.19)	<b>110</b> (3.68)	<b>1.43</b>	<b>1.13</b>	<b>4</b>		
心理學	77	110	1.43	1.13		7	9

在資料使用數量上，歧異程度甚高。

在學科部分，如果依據資料使用數量來看，六大學科正好也位居前六名，其順序為經濟、農經、政治、公行國關、社會，與教育。經濟學的使用量幾乎是排名第二的農經的兩倍（1001：503），更為排名第六的教育學的4倍有餘（1001：243），顯示了該學科在資料再用上的強度。而如以論文平均資料使用量來看，部分學科如傳播學、社工等，排名會比六大學科高，但那是因為該學科的論文篇數少，只要少量論文使用較多筆資料，就會提高學科平均使用量。如果單獨比較六大學科的平均使用量，由高至低順序正好也與數量排名相同。換言之，在學科層次上，資料使用量最高者，依序就是經濟、農經、政治、公行國關、社會，與教育。唯其中農經與政治學標準差數值較大，顯示這兩個學科在資料使用量上歧異程度也較大。

從累計百分比來觀察（參見附錄2），可發現在整體樣本中，約有六成（60.58%）的再用論文僅使用1筆資料；使用1-2筆者累計略高過整體的3/4（77.02%）；而使用1-4筆資料者，累計即超過論文九成（91.64%）。但進一步觀察五個學門時，可發現在累計分布上有明顯的差距。經濟學門中僅用1筆資料的論文低於五成（44.56%），政治學則僅逼近六成（58.06%），也就是說，在這兩個學門中，使用多筆資料的再用論文相當常見，甚至有單篇論文可同時使用20餘筆資料。相對而言，其他三個學門都以使用1筆資料為大宗，數量都佔學門論文的3/4以上（社會學最低，為75.72%，教育學門高達84.51%），換言之，在這三個學門中，使用1筆以上資料的論文都屬小眾。

而觀察六大學科累計百分比分布（參見附錄3），可發現與學科與所屬學門的分布狀況非常接近；而且，屬同一學門的學科，累計分布的趨勢也相當接近。首先，同屬經濟學門的經濟與農經，僅用1筆資料的再用論文比例都未過半（42.28%與49.01%）；而將近八成的再用論文使用資料都是介於1-3筆（77.72%與80.20%）；稍逾九成的論文使用1-5筆（92.66%與91.58%）。再來，同屬政治學門的政治與公行國關，僅用1筆資料者，均是高於五成而未達六成（57.75%與58.44%）；使用1-2筆者均佔再用論文約莫3/4（78.07%與75.32%）；而同樣的，使用1-4筆者即逾九成（90.91%與92.86%）。這個結果顯示，在經濟與政治學門中，其下屬學科的資料再用行為是相當類似的。

而社會學與教育學這兩個學科，因分別是社會學門與教育學門下的資料再用大宗，學門下的其他學科因再用論文稀少，因此，學科的分布趨勢幾乎與學門完全相當。在社會學中，僅用1筆資料者即逾再用論文的3/4（76.92%）；而教育學中僅用1筆資料者，就已超過再用論文的八成五（86.2%）。

本研究另觀察樣本整體及學門／學科層次的資料再用年均分布，目的有二：其一，觀察在近年資料分享的呼籲下，社會科學的資料再用是否有隨之增加；其二，不同學門或學科之間，資料再用是否有獨特的成長變化。圖3與圖4

分別顯示整體樣本及學門／學科領域中，15年區間內再用論文的資料使用筆數年均分布狀況。

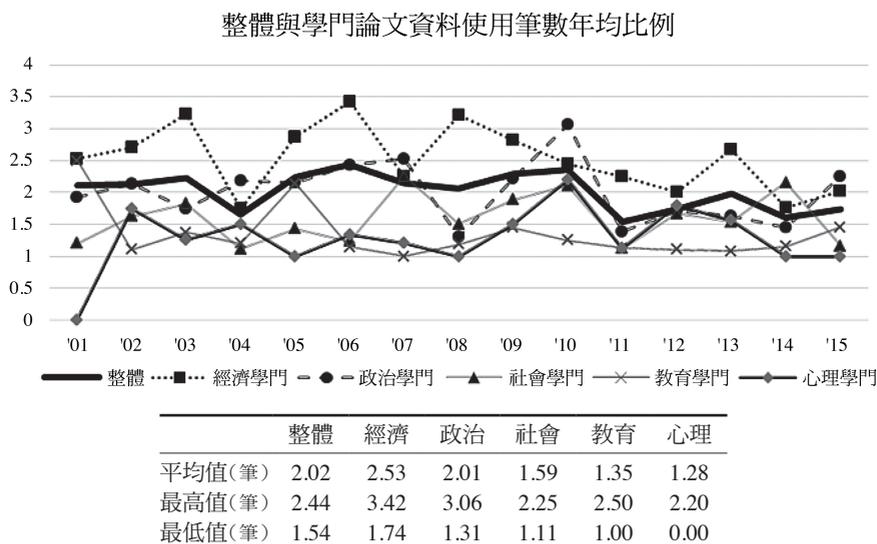


圖3 整體樣本與各學門論文資料使用量年均分布曲線圖

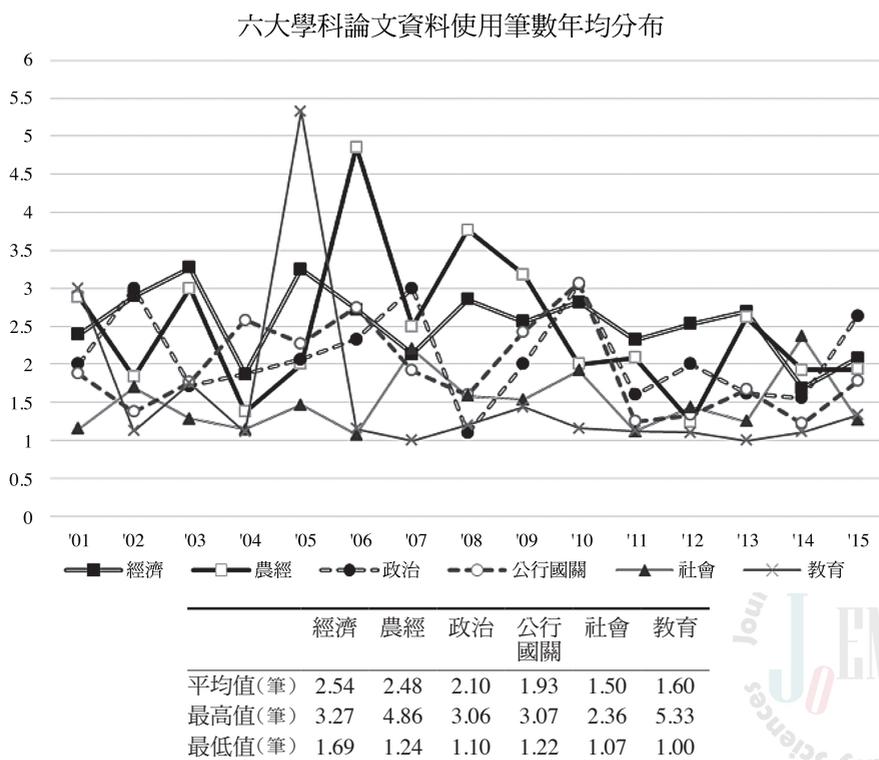


圖4 六大學科論文使用資料量年均分布曲線圖

首先，就整體而言，年均使用資料量起伏不大，最高值為2.44筆，最低值為1.54筆，與平均值相差僅0.42-0.48筆。在學門部分，15年間的年均分布或有起伏，但同樣未見有任一學門出現明顯持衡的上升或下降趨勢。六大學科的分布亦然，若干學科如農經與教育學，曾於特定年間出現使用多筆資料的論文，因而拉高該年年均筆數，但整體來看，也無任何特定學科出現持衡的上升或下降趨勢，顯示在本研究分析的時限內，似無外在趨力在系統性地鼓勵或貶抑同篇論文中多筆資料的複合利用。

### (三) 被引資料的來源分析

本研究依被引資料的蒐集者或彙編者將資料來源區分成政府機關、學術機構、民間單位、與個別研究四種。表3顯示：在整體樣本中，來源為政府機關的被引資料佔比最高，超過一半(53.88%)；學術機構次之，佔比近兩成(18.70%)；來自民間單位與個別研究的資料比例都低，尤其以先前個別研究進行衍生的原創分析，在整體樣本中不到一成(8.53%)。

比較五個學門的被引資料來源，除去資料再用數量甚低的心理學門，可發現政府機關是其他四個學門所用資料來源之最，但在各學門的佔比上差距甚大，經濟學門高達66.56%，社會學門也超過一半(52.47%)，相對而言，政治與教育學門的使用量僅佔三至四成之間。學術機構在政治、社會，與教育三個學門的被引資料中，佔比約1/3(政治學門：34.44%；社會學門：32.47%)至1/4(教育學門：26.26%)，但在經濟學門就佔比極低(僅4.99%)。來自民間單位與

表3 被引資料來源分布(2001-2015)

學門/學科	資料蒐集者 彙編者		政府機關		學術機構		民間單位		個別研究		無法辨識	
	數量	%	數量	%	數量	%	數量	%	數量	%	數量	%
<b>整體 (N=2,990)</b>	<b>1,611</b>	<b>53.88</b>	<b>559</b>	<b>18.70</b>	<b>458</b>	<b>15.32</b>	<b>255</b>	<b>8.53</b>	<b>107</b>	<b>3.58</b>		
<b>經濟學門 (n=1,504)</b>	<b>1,001</b>	<b>66.56</b>	<b>75</b>	<b>4.99</b>	<b>302</b>	<b>20.08</b>	<b>58</b>	<b>3.86</b>	<b>68</b>	<b>4.52</b>		
經濟學 (n=1,001)	666	66.53	51	5.09	224	22.38	24	2.40	36	3.60		
農業經濟 (n=503)	335	66.60	24	4.77	78	15.51	34	6.76	32	6.36		
<b>政治學門 (n=694)</b>	<b>255</b>	<b>36.74</b>	<b>239</b>	<b>34.44</b>	<b>62</b>	<b>8.93</b>	<b>117</b>	<b>16.86</b>	<b>21</b>	<b>3.03</b>		
政治學 (n=389)	119	30.59	132	33.93	41	10.54	87	22.37	10	2.57		
公行國關 (n=305)	136	44.59	107	35.08	21	6.89	30	9.84	11	3.61		
<b>社會學門 (n=385)</b>	<b>202</b>	<b>52.47</b>	<b>125</b>	<b>32.47</b>	<b>28</b>	<b>7.27</b>	<b>23</b>	<b>5.97</b>	<b>7</b>	<b>1.82</b>		
社會學 (n=291)	138	47.42	114	39.18	14	4.81	19	6.53	6	2.06		
社會工作 (n=57)	51	89.47	3	5.26	1	1.75	1	1.75	1	1.75		
傳播學 (n=37)	13	35.14	8	21.62	13	35.14	3	8.11	0	0.00		
<b>教育學門 (n=297)</b>	<b>118</b>	<b>39.73</b>	<b>78</b>	<b>26.26</b>	<b>49</b>	<b>16.50</b>	<b>43</b>	<b>14.48</b>	<b>9</b>	<b>3.03</b>		
教育學 (n=243)	93	38.27	77	31.69	27	11.11	43	17.70	3	1.23		
體育學 (n=45)	21	46.67	0	0.00	18	40.00	0	0.00	6	13.33		
圖書資訊學 (n=9)	4	44.44	1	11.11	4	44.44	0	0.00	0	0.00		
<b>心理學門 (n=110)</b>	<b>35</b>	<b>31.82</b>	<b>42</b>	<b>38.18</b>	<b>17</b>	<b>15.45</b>	<b>14</b>	<b>12.73</b>	<b>2</b>	<b>1.82</b>		
心理學 (n=110)	35	31.82	42	38.18	17	15.45	14	12.73	2	1.82		

個別研究的資料，學門使用比例均低，多數低於兩成。

另一個值得注意的狀況是：對這四個資料再用較發達的學門而言，使用量都集中在兩個主要來源，將最高的兩個來源相加，即可佔全部被引資料的六至八成以上。換言之，每個學門均有其倚重的兩大特定來源，唯學門間兩大來源的比例有差距：經濟學門以政府機關及民間單位的資料為主；政治、社會，及教育學門則以政府機關及學術機構資料為主。但教育學相對於其他三個學門，資料來源的分布比較沒有那麼集中。卡方檢定顯示五個學門的資料來源分布（不含無法辨識者）具顯著差異，學門別與資料來源呈強度相關（ $\chi^2=611.404$ ,  $df=8$ ,  $p<0.001$ , Cramer's  $V=.322$ ）。

就六大學科的狀況來看，分布趨勢亦大致與學科所屬學門相近。經濟與農經均高度仰賴政府機關的資料，民間單位者次之，且比例已相當低；政治學、公行國關、社會學、教育學均較仰賴政府機關及學術機構的資料，且兩個來源的分布相對較均勻，兩者相加可達資料的六至八成。但也有若干歧異出現，例如，同屬政治學門的政治學與公行國關，在個別研究資料的使用比例上就出現明顯差距；教育學使用學術機構與個別研究的資料比例，也略高於教育學門整體。卡方檢定顯示：六大學科的資料來源分布呈顯著差異（ $\chi^2=666.707$ ,  $df=15$ ,  $p<0.001$ , Cramer's  $V=.290$ ）。更甚者，同屬經濟學門的經濟學與農經，資料來源分布也具顯著差異（ $\chi^2=24.428$ ,  $df=3$ ,  $p<0.001$ , Cramer's  $V=.130$ ）；同列政治學門下的政治學與公行國關亦然（ $\chi^2=26.675$ ,  $df=3$ ,  $p<0.001$ , Cramer's  $V=.203$ ）。上述結果顯示，學科別與資料來源呈中度相關。

#### （四）被引資料類型分析

本研究將資料類型區分為「系列調查」、「業務資料」及「一次性資料」（定義參見前節說明）。表4顯示，在整體樣本中，業務資料的使用量超過全部資料的一半（54.31%）；系列調查的使用略高於三成（32.31%）；一次性資料的使用僅佔一成左右（11.97%）。

就學門來看，經濟學門使用的資料類型與其他學門大不相同，以業務資料為大宗，使用比例高達3/4（74.87%）；相對而言，其他學門的業務資料使用量，都約佔1/3而已。社會、教育，與心理學門使用系列調查資料的比例則比較高，都超過一半以上。政治學門使用三種資料類型的比例則相對均勻，三種資料使用量的比值約為1.38：1.56：1。卡方檢定顯示五個學門之間，資料類型的分布具顯著差異，學門別與資料類型呈強度相關（ $\chi^2=611.404$ ,  $df=8$ ,  $p<0.001$ , Cramer's  $V=.322$ ）。

就六大學科的資料類型分布來看，同屬經濟學門的經濟學與農經，都是以業務資料居多，且都佔整體資料3/4上下。但同屬政治學門下的政治學與公行國關，資料類型的使用分布差異就比較大，在三種資料類型上，使用比例

表4 資料類型分布(2001-2015)

學門學科	業務資料		系列調查		一次性者		無法辨識	
	數量	%	數量	%	數量	%	數量	%
<b>整體 (N=2,990)</b>	<b>1,624</b>	<b>54.31</b>	<b>966</b>	<b>32.31</b>	<b>358</b>	<b>11.97</b>	<b>42</b>	<b>1.40</b>
<b>經濟學門 (n=1,504)</b>	<b>1,126</b>	<b>74.87</b>	<b>266</b>	<b>17.69</b>	<b>85</b>	<b>5.65</b>	<b>27</b>	<b>1.80</b>
經濟學 (n=1,001)	739	73.83	197	19.68	42	4.20	23	2.30
農業經濟 (n=503)	387	76.94	69	13.72	43	8.55	4	0.80
<b>政治學門 (n=694)</b>	<b>241</b>	<b>34.73</b>	<b>273</b>	<b>39.34</b>	<b>174</b>	<b>25.07</b>	<b>6</b>	<b>0.86</b>
政治學 (n=389)	95	24.42	174	44.73	115	29.56	5	1.29
公行國關 (n=305)	146	47.87	99	32.46	59	19.34	1	0.33
<b>社會學門 (n=385)</b>	<b>121</b>	<b>31.43</b>	<b>208</b>	<b>54.03</b>	<b>54</b>	<b>14.03</b>	<b>2</b>	<b>0.52</b>
社會學 (n=291)	74	25.43	172	59.11	43	14.78	2	0.69
社會工作 (n=57)	29	50.88	20	35.09	8	14.04	0	0.00
傳播學 (n=37)	18	48.65	16	43.24	3	8.11	0	0.00
<b>教育學門 (n=297)</b>	<b>100</b>	<b>33.67</b>	<b>161</b>	<b>54.21</b>	<b>32</b>	<b>10.77</b>	<b>4</b>	<b>1.35</b>
教育學 (n=243)	59	24.28	151	62.14	30	12.35	3	1.23
體育學 (n=45)	35	77.78	8	17.78	1	2.22	1	2.22
圖書資訊學 (n=9)	6	66.67	2	22.22	1	11.11	0	0.00
<b>心理學門 (n=110)</b>	<b>36</b>	<b>32.73</b>	<b>58</b>	<b>52.73</b>	<b>13</b>	<b>11.82</b>	<b>3</b>	<b>2.73</b>
心理學 (n=110)	36	32.73	58	52.73	13	11.82	3	2.73

差距可達10%-20%以上。社會學與教育學的資料類型分布比較相近，都是系列調查佔一半以上，業務資料佔1/4左右，一次性資料略高於一成。卡方檢定顯示六大學科之間，資料類型分布具顯著差異，學科別與資料類型呈強度相關( $\chi^2=693.443$ ,  $df=10$ ,  $p<0.001$ , Cramer's  $V=.359$ )。而同屬經濟學門的經濟學與農經，資料類型分布也具顯著差異，學科別與資料類型呈中度相關( $\chi^2=18.218$ ,  $df=2$ ,  $p<0.001$ , Cramer's  $V=.111$ )；同屬政治學門的政治學與公行國關亦然( $\chi^2=40.667$ ,  $df=2$ ,  $p<0.001$ , Cramer's  $V=.243$ )。

## 五、結論與建議

本研究調查15年間台灣社會科學學門領域內的量化資料再用情況，在期刊來源上，採取TSSCI收錄期刊為分析對象，足以代表台灣社會科學重要學科領域的研究行為。有別於前人以抽樣、推估，或從資料集來反尋再用論文的作法，本研究採逐篇檢視、全面清查的方式，在呈現資料再用的全貌上，證據基礎應較堅實，準確程度也較高。

如以本文所分析的五大學門為社會科學代表，則台灣社會科學學界透過資料再用來產出實徵論文的比例，低於1/5(佔整體實徵論文17.38%)，且15年間比例大致持衡，年均介於15%-22%，未見明顯增長或消退趨勢，與科學社群近年在資料再用比例出現明顯成長，大不相同(Piwovar & Vision, 2013; Xia &

Liu, 2013; Zhang et al., 2011)。資料再用對於社會科學整體實徵研究的貢獻，規模可能大致就是如此。

資料再用在不同的社會學科領域之間，強度不同。觀察五大學門及其下屬學科，於資料再用比例上具有差距。本文認為這些差異應可歸因於學科本質差異，具體來說，經濟領域的資料再用高度普及，資料再用甚至可說是實徵研究的主流現象；而政治研究領域的資料再用程度也高，這是因為這兩個領域的主題關懷較常是宏觀現象，因此取用外部資料就可能是必要且唯一的研究途徑，相對而言，社會、教育，與心理領域在研究主題上，雖也可能分析宏觀或長期趨勢，因而取用外部資料，但整體來說，這些領域學者在主題上，從宏觀至微觀，切入層次相對多元，研究者能獨立蒐集資料的機會也比較大，因此資料再用程度較低。

林奇秀與賴璟毅(2017)的訪談結果也間接呼應上述觀點：資料再用的主要理由，最常是因為研究主題需要宏觀層級、大規模，且個人無力能及的資料，而經濟與政治的實徵研究即以這類為大宗。此外，歐美國家在1960年代即成立的社會科學資料典藏中心，名稱也突顯了經濟與政治領域對外部資料的強烈需求，例如當時英國的Social and Economic Archive Committee即以經濟資料為大宗；美國當時所成立的Inter-university Consortium for Political Research(即ICPSR前身)，與私人民意調查機構Roper Public Opinion Research Center，均以政治與公共事務資料為大宗(Nasatir, 1967)。由此即可看出，經濟與政治這兩大領域對外部資料的高度仰賴。

相對而言，台灣學界在心理學研究上，資料再用偏低，則與國外研究發現有出入。國外研究並無對心理學整體資料再用的調查結果，但Fear(2013)在探討社會科學第三方資料再用時，發現資料再用主題排名前五名之中，包含有「醫學與心理類」(排名第二)與「心理學與社會科學類」(排名第五)。雖然其分析基礎與本文完全不同，但單就第三方再用來看，心理學顯然具有相當程度的資料再用，是個再用程度不低的領域。但在台灣，心理學無論在整體資料再用或取用個別研究資料來進行衍生分析，數量都非常少。而且，如前文指出，本研究的心理學期刊中有四份是教育心理學期刊，佔了本研究心理學論文樣本近七成(68.83%)，更凸顯出心理學資料再用的罕見。林奇秀與賴璟毅(2017)所訪談的學者曾表示：台灣的心理學界並不鼓勵資料再用。這個說法與本研究的分析結果是相符的。

本文亦從資料使用筆數、資料類型，與資料來源上辨析出不同學科領域的資料再用特徵。首先，經濟與政治學門不但是資料再用的大宗，而且平均來說，每篇論文使用的資料筆數也是較高的領域，一篇論文同時多筆資料相當常見；反之，社會、教育，與心理領域中，均以使用一筆資料為主。其次，在

資料來源上，政府機關是所有領域都最倚重的資料來源，但經濟與社會領域比其他領域都更加仰賴政府資料；相對而言，政治、教育，與心理對學術機構的資料，也有極高的使用，但經濟卻罕有使用學術機構資料。再者，在資料類型方面，經濟領域再度與眾不同，高度仰賴業務資料；相對而言，社會、教育、與心理則較仰賴系列調查；政治在業務資料與系列調查的使用量則約略相當。

然而，上述領域亦有共同之處：無論是對來源為個別研究或資料類型為一次性資料者，在每個領域的使用比例都非常低。這個結果與科學領域的資料再用研究發現大不相同。從既有文獻可得知，基因的資料再用幾乎均是先前研究的資料再用，天文或其他科學領域研究也有相當數量的資料再用是使用先前的個別研究數據 (Piwowar et al., 2011; Piwowar & Vision, 2013; Xia & Liu, 2013; Zhang, 2011)，但本文發現個別研究數據在台灣社會科學領域中，整體被用的比例非常低，不足全部資料集的一成；而一次性資料的使用也僅略高於一成。從這個結果來看，社會科學所需要的資料，主要是組織性、長期性的大規模資料，無論足以反映公私部門運行狀況的業務數據，或呈現社會現象逐年變遷的調查統計。林奇秀與賴環毅 (2017) 對社會科學學者的訪談也顯示：個別研究資料有其特定問題意識，且有其特定產出時空情境，用在社會科學的衍生分析上，效用比較有限。綜合此次的研究結果，本文建議：社會科學的資料採用服務，與其汲汲於向學者徵集個別研究產出的資料，或許更應著力於蒐集、彙整、媒合來自公私部門的大規模、長期性數據資料，倘能進一步建立跨部會，甚或跨政府與民間單位的資料服務機制，讓資料資訊更透明，並降低資料取用障礙，如此應更能有效促成社會科學的資料再用，發揚既有資料的存在價值。

但本研究也具有一定的分析侷限，例如，從活化資料、提高再用的角度來看，社會科學資料的調查週期或發佈時間對再用的影響，是一個非常值得關注的議題，然而，本文的研究方法並無法回答這個問題，主要是目前學界在資料引註上甚乏明確統一的引文規範，不同作者提供的資料描述內容及其交代方式，歧異很大，我們或可從論文內文或原始資料集來判斷資料數據起迄時間，但不見得都能獲知其發佈時間與調查週期，也因此，這個攸關資料傳播與再用結構的問題，目前無法透過內容分析來觀察，有賴未來研究以其他方法探索之。此外，本文顯示不同學科在資料來源與資料類型上具有差異，可見資料再用行為具有領域性，然而，具體而言，資料再用行為受到哪些學科領域特質影響、造成哪些領域性的變異、這些變異對資料分享與再用帶來哪些促成或箝制的效果等，目前均尚屬未知，有待未來研究進一步探討。

## 誌 謝

本文為科技部專題研究計畫 (MOST 105-2410-H-002-120-MY2) 之部分研究成果。

## 參考文獻

- 林奇秀、賴璟毅 (2017)。臺灣社會科學學者資料再用行為之研究。圖書資訊學研究, 11(2), 95-138。
- 科技部人文社會科學研究中心 (2015)。TSSCI收錄期刊名單。檢自<http://www.hss.ntu.edu.tw/model.aspx?no=67>
- 陳雪華 (2012)。導論。在陳雪華、陳光華 (編著), *e-Research* 學術圖書館創新服務 (頁 1-13)。台北市: 國立臺灣大學圖書館。
- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059-1078. doi:10.1002/asi.22634
- Broom, A., Cheshire, L., & Emmison, M. (2009). Qualitative researchers' understandings of their practice and the implications for data archiving and sharing. *Sciology*, 43(6), 1163-1180. doi:10.1177/0038038509345704
- Chao, T. C. (2012, February). *Exploring the rhythms of scientific data use*. In *Proceedings of the 2012 iConference* (pp. 129-135). New York, NY: ACM. doi:10.1145/2132176.2132193
- Data Curation Center. (n.d.). What is data curation? Retrieved from <http://www.dcc.ac.uk/digital-curation/what-digital-curation>
- Data Information Specialist Committee - UK. (2010). Questions & answers. Retrieved from <http://www.disc-uk.org/qanda.html>
- Fear, K. M. (2013). *Measuring and anticipating the impact of data reuse* (Unpublished doctoral dissertation). Information and Library Studies, University of Michigan, Ann Arbor, MI.
- Henderson, M. E., & Knott, T. L. (2015). Starting a research data management program based in a university library. *Medical Reference Services Quarterly*, 34(1), 47-59. doi:10.1080/02763869.2015.986783
- Jeng, W. (2017). *Qualitative data sharing practices in social sciences* (Unpublished doctoral dissertation). School of Information Sciences, University of Pittsburgh, Pittsburgh, PA.
- Kim, J. (2013). Data sharing and its implications for academic libraries. *New Library World*, 114(11/12), 494-506. doi:10.1108/NLW-06-2013-0051
- Konkiel, S. (2013). Tracking citation and altmetrics for research data: Challenges and opportunities. *Bulletin of the American Society for Information Science and Technology*, 39(6), 27-32. doi:10.1002/bult.2013.1720390610
- Kruse, F., & Thestrup, J. B. (2014). Research libraries' new role in research data management, current trends and visions in Denmark. *LIBER Quarterly*, 23(4), 310-335. doi:10.18352/lq.9173
- MacMillan, D. (2014). Data sharing and discovery: What librarians need to know. *Journal of Academic Librarianship*, 40(5), 541-549. doi:10.1016/j.acalib.2014.06.011

- Nasatir, D. (1967). Social science data libraries. *The American Sociologist*, 2(4), 207-212.
- National Research Council. (1999). *A question of balance: Private right and the public interest in scientific and technical databases*. Washington, DC: National Academies Press.
- Piwowar, H. A. (2008). Proposed foundations for evaluation data sharing and reuse in the biomedical literature. *Bulletin of IEEE Technical Committee on Digital Libraries*, 4(2). Retrieved from <http://www.ieee-tcdl.org/Bulletin/v4n2/piwowar/piwowar.html>
- Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, e175. doi:10.7717/peerj.175
- Piwowar, H. A., Carlson, J. D., & Vision, T. J. (2011). Beginning to track 1000 datasets from public repositories into the published literature. *Proceedings of the American Society for Information Science and Technology*, 48(1), 1-4. doi:10.1002/meet.2011.14504801337
- Robinson-García, N., Jiménez-Contreras, E., & Torres-Salinas, D. (2016). Analyzing data citation practices using the data citation index. *Journal of the Association for Information Science and Technology*, 67(12), 2964-2975. doi:10.1002/asi.23529
- Sechrest, L., & Sidani, S. (1995). Quantitative and qualitative methods: Is there an alternative? *Evaluation and Program Planning*, 18(1), 77-87. doi:10.1016/0149-7189(94)00051-X
- Van den Berg, H. (2005). Reanalyzing qualitative interviews from different angles: The risk of decontextualization and other problems of sharing qualitative data. *Forum: Qualitative Social Research*, 6(1). doi:10.17169/fqs-6.1.499
- Xia, J., & Liu, Y. (2013). Usage patterns of open genomic data. *College & Research Library*, 74(2), 195-207. doi:10.5860/crl-324
- Zhang, J. (2011). *Data use and access behavior eScience — Exploring data practices in the new data-intensive science paradigm* (Unpublished doctoral dissertation). Department of Information Science, Drexel University, Philadelphia, PA.
- Zhang, J., Vogeley, M. S., & Chen, C. (2011). Scientometrics of big science: A case study of research in the Sloan Digital Sky Survey. *Scientometrics*, 86(1), 1-14. doi:10.1007/s11192-010-0318-1

## 附錄一 本研究期刊樣本之學門/學科列表

社會學門(11份)	政治學門(11份)	教育學門(21份)	經濟學門(7份)	心理學門(7份)
<b>社會學(5份)</b> 人口學刊* 文化研究☆(2005~) 台灣社會學 臺灣社會學刊 調查研究：方法與應用	<b>政治學(6份)</b> 東吳政治學報 政治科學論叢 政治與社會哲學 評論☆(2002~) 政治學報 臺灣民主季刊 ☆(2004~) 台灣政治學刊	<b>教育學(17份)</b> 科學教育學刊 特殊教育研究學 刊 特殊教育學報 教育政策論壇 教育研究集刊 教育科學研究期 刊★(2009~) 教育實踐與研究 ★(2007~) 教育學刊 教育學報 教科書研究 ☆(2008~) 華語文教學研究 ☆(2004~) 當代教育研究季 刊★(2005~) 臺東大學教育學 報★(2003~) 臺灣教育社會學 研究 數位學習科技期 刊☆(2008~) 課程與教學季刊 藝術教育研究	<b>經濟學(4份)</b> 經濟研究 經濟論文 經濟論文叢刊 臺灣經濟預測與 政策* <b>農業經濟(3份)</b> 農業經濟叢刊 農業與經濟 應用經濟論叢 ★(2008~)	<b>心理學(7份)</b> 中華心理衛生學 刊* 中華心理學刊 中華輔導與諮商 學報★(2008~) 本土心理學研究 教育心理學報 教育與心理研究* 測驗學刊* ★(2003~)
<b>社工學(2份)</b> 社會政策與社會 工作學刊 臺大社會工作學 刊	<b>公行國關(5份)</b> 公共行政學報 行政暨政策學報 問題與研究季刊 遠景基金會季刊 ★(2002~) 選舉研究	<b>體育學(2份)</b> 大專體育學刊 體育學報		
<b>傳播學(4份)</b> 新聞學研究 中華傳播學刊 ☆(2002~) 傳播與社會學刊 ☆(2006~) 傳播研究與實踐 ☆(2011~)		<b>圖資學(2份)</b> 教育資料與圖書 館學 圖書資訊學研究 ★(2006~)		

註：\*為原列於TSSCI「綜合類」之期刊

☆為2001年後才出現的新刊(括號為發刊年份)

★為期刊更名者(括號為更名後首次發刊年份)



附錄二 整體與學門再用論文資料使用筆數累計分布

學門 被引 資料數	整體			經濟學門			政治學門			社會學門			教育學門			心理學門					
	數量	%	累計 數量	數量	%	累計 數量	數量	%	累計 數量	數量	%	累計 數量	數量	%	累計 數量	數量	%	累計 數量			
1	899	60.58	899	266	44.56	266	198	58.06	198	184	75.72	184	75.72	191	84.51	191	84.51	60	77.92	60	77.92
2	244	16.44	1,143	127	21.27	393	64	18.77	262	25	10.29	209	86.01	19	8.41	210	92.92	9	11.69	69	89.61
3	138	9.30	1,281	76	12.73	469	30	8.80	292	17	7.00	226	93.00	10	4.42	220	97.35	5	6.49	74	96.10
4	79	5.32	1,360	49	8.21	518	21	6.16	313	4	1.65	230	94.65	3	1.33	223	98.67	2	2.60	76	98.70
5	56	3.77	1,416	33	5.53	551	16	4.69	329	6	2.47	236	97.12	1	0.44	224	99.12	0	0.00	76	98.70
6	21	1.42	1,437	16	2.68	567	4	1.17	333	1	0.41	237	97.53	0	0.00	224	99.12	0	0.00	76	98.70
7	16	1.08	1,453	9	1.51	576	3	0.88	336	3	1.23	240	98.77	1	0.44	225	99.56	0	0.00	76	98.70
8	11	0.74	1,464	8	1.34	584	2	0.59	338	1	0.41	241	99.18	0	0.00	225	99.56	0	0.00	76	98.70
9	2	0.13	1,466	0	0.00	584	0	0.00	338	1	0.41	242	99.59	0	0.00	225	99.56	0	0.00	76	98.70
10	6	0.40	1,472	4	0.67	588	1	0.29	339	1	0.41	243	100.00	0	0.00	225	99.56	1	1.30	77	100.00
11	3	0.20	1,475	3	0.50	591	0	0.00	339	0	0.00	243	100.00	0	0.00	225	99.56	0	0.00	76	98.70
12	2	0.13	1,477	2	0.34	593	0	0.00	339	0	0.00	243	100.00	0	0.00	225	99.56	0	0.00	76	98.70
14	2	0.13	1,479	1	0.17	594	0	0.00	339	0	0.00	243	100.00	1	0.44	226	100.00	0	0.00	76	98.70
15	1	0.07	1,480	1	0.17	595	0	0.00	339	0	0.00	243	100.00	0	0.00	225	99.56	0	0.00	76	98.70
18	1	0.07	1,481	0	0.00	595	1	0.29	340	1	0.41	243	100.00	0	0.00	225	99.56	0	0.00	76	98.70
19	1	0.07	1,482	1	0.17	596	0	0.00	340	0	0.00	243	100.00	0	0.00	225	99.56	0	0.00	76	98.70
25	1	0.07	1,483	0	0.00	596	1	0.29	341	1	0.41	243	100.00	0	0.00	225	99.56	0	0.00	76	98.70
27	1	0.07	1,484	1	0.17	597	0	0.00	341	0	0.00	243	100.00	0	0.00	225	99.56	0	0.00	76	98.70
總計	1,484	100.00		597	100.00		341	100.00		243	100.00		226	100.00		77	100.00				

## 附錄三 六大學科再用論文資料使用筆數累計分布

學門 被引 資料數	經濟		農經		政治		公行國關		社會		教育	
	數量	累計 數量 %	數量	累計 數量 %	數量	累計 數量 %	數量	累計 數量 %	數量	累計 數量 %	數量	累計 數量 %
1	167	42.28	99	49.01	108	57.75	90	58.44	150	76.92	160	86.02
2	88	22.28	39	19.31	38	20.32	26	16.88	19	9.74	15	8.06
3	52	13.16	24	11.88	13	6.95	17	11.04	16	8.21	5	2.69
4	39	9.87	10	4.95	11	5.88	10	6.49	3	1.54	3	1.61
5	20	5.06	13	6.44	8	4.28	8	5.19	4	2.05	1	0.54
6	10	2.53	6	2.97	4	2.14	0	0.00	0	0.00	0	0.00
7	6	1.52	3	1.49	1	0.53	2	1.30	2	1.03	1	0.54
8	3	0.76	5	2.48	2	1.07	0	0.00	0	0.00	0	0.00
9	0	0.00	0	0.00	0	0.00	0	0.00	1	0.51	0	0.00
10	3	0.76	1	0.50	1	0.53	0	0.00	0	0.00	0	0.00
11	3	0.76	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
12	2	0.51	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
14	1	0.25	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
15	0	0.00	1	0.50	0	0.00	0	0.00	0	0.00	1	0.54
18	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
19	1	0.25	0	0.00	1	0.53	1	0.65	1	0.54	1	0.54
25			0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
27			1	0.50	1	0.53	0	0.00	0	0.00	0	0.00
總計	395	100.00	202	100.00	187	100.00	154	100.00	195	100.00	186	100.00



# The Reuse of Quantitative Data in Social Sciences in Taiwan: 2001-2015

Chi-Shiou Lin<sup>a\*</sup> Ching-Yi Lai<sup>b</sup>

Research Article

## Abstract

*This study explored the reuse of existing quantitative data in original social sciences research in Taiwan. Using the 2015 TSSCI List as the basis of journal selection, this study focused on 57 journals listed under the subject divisions of economics, political science, sociology, education, and psychology. Journal issues published between 2001 and 2015 were manually scanned to identify data reuse papers. The characteristics of the reuse papers and the cited data were recorded for the subsequent subject division-level and subject discipline-level analyses. A total of 1,484 reuse papers were identified, which accounted for 17.38% of the total empirical study papers. Among the five subject divisions, economics and political science had the highest percentages of data reuse, while psychology had the lowest. Those reuse papers together cited 2,990 datasets. Most of the datasets were used in economics and political science papers. Further, these two subject divisions had noticeably larger proportions of papers that use more than one dataset. In contrast, papers using only one dataset were the majority for the rest. In regards to data source and data type, datasets originated from government agencies as well as data generated from business operations accounted for the majority of the cited data, but significant differences existed among those subjects. As opposed to the frequently used business and series survey data, data generated from independent research projects as well as data that were non-continuous and one-time in nature were rarely reused in social sciences research. Based on the study findings, it is recommended that data services for social sciences research should focus more on large-scaled continuous data generated from governments and research institutions. A service mechanism that bridges users and data providers from public and private sectors would also enhance data reuse and increase the value of existing data.*

**Keywords:** Data reuse, Data citation, Quantitative data, Social sciences

## SUMMARY

The rise of e-science and data curation in the recent decades has prompted the sharing and reuse of research data in sciences. However, investigations on how and to what extent the existing data have been used for derivative, original

<sup>a</sup> Associate Professor, Department of Library & Information Science, National Taiwan University, Taipei, Taiwan

<sup>b</sup> Graduate Student, Department of Library & Information Science, National Taiwan University, Taipei, Taiwan

\* Principal author for all correspondence. E-mail: chishioulin@ntu.edu.tw

research are still rare. Among the limited number of research, most has focused on data reuse in science disciplines, particularly, in genetics and astronomy. Rarely has the existing research examined data reuse in social sciences.

Further, due to the lack of standardized data citation practices in current academic writings, most of the existing data reuse studies have identified data reuse papers based on the presence of data registry keys (e.g., the DOIs of datasets of a particular data repository) in the references and/or the main texts of research papers. While this method works well for some scientific disciplines that have data sources that are well archived by one or two data repositories, it is less effective for identifying data reuse papers in social sciences as the potential data sources are more disperse and sporadically distributed. Moreover, not all of the datasets in social science papers are deposited in data repositories and there might not be registry keys for those datasets that have actually been reused. For social science research outside of the English-speaking world, relying on data registry keys is even more impractical as most of the reused datasets may have never been registered in international data repositories or data citation indexes.

This study thus adopted another approach. The authors manually scanned the empirical research papers published within a body of social science journals to identify all papers that have empirical analysis which were based wholly or partially on existing datasets. With the papers drawn from a time span of 15 years (2001-2015), this study sought to answer the following questions:

1. To what extent have data reuse papers accounted for the social science research papers? Were there observable growths in data reuse in social sciences over the 15 years?
2. How many datasets have been used in the data reuse papers? Were there observable growths in dataset usage?
3. What were the major sources of data for the social sciences disciplines? Where there significant differences in data sources among various social sciences disciplines?
4. What were the major data types for social science disciplines? Were there significant differences in data types among various social sciences disciplines?

This study employed content analysis on the papers published by 57 journals that were included in the 2015 journal list from the Taiwan Social Sciences Citation Index (TSSCI). The 57 journals together represented five larger subject divisions (i.e., economics, political sciences, social sciences, education, and psychology) or eleven smaller subject disciplines. Each paper published between 2001 and 2015 was examined to ascertain if it constituted a data reuse paper. For each data reuse paper, the data reuse characteristics as well as the characteristics of the used datasets were systematically recorded for later analysis.

Based on the analyses, it was found that data reuse papers accounted for 11.99% of the entire sample of research papers (including empirical and non-empirical research papers) and 17.38% for the empirical study papers. The subject division of economics constituted the largest user group of existing data in empirical research; it claimed 40.23% of the total data reuse papers (Table 1). However, the analysis of chronological distributions of the data reuse papers showed no obvious increase or decrease of data reuse in the entire sample or by subject division/discipline.

**Table 1 The Distribution of Data Reuse Papers in Taiwan Social Sciences Research, 2001-2015**

Subject Division/ Subject Discipline	Number of research papers	Number of empirical papers	No. of the data reuse papers (% for the total reuse papers)	% - total research papers	% - total empirical papers	Rank 1*	Rank 2*
<b>Total</b>	<b>12,381</b>	<b>8,541</b>	<b>1,484(100.00)</b>	<b>11.99</b>	<b>17.38</b>		
<b>Economics Div.</b>	<b>1,125</b>	<b>750</b>	<b>597(40.23)</b>	<b>53.07</b>	<b>79.60</b>		
Economics	764	469	395	51.70	84.22	1	1
Agri. Eco.	361	281	202	55.96	71.89	2	2
<b>Political Sci. Div.</b>	<b>2,265</b>	<b>836</b>	<b>341(26.62)</b>	<b>15.06</b>	<b>40.79</b>		
Political Science	1,216	442	187	15.38	42.31	4	4
Public Admin. & Int'l Affairs	1,049	394	154	14.68	39.09	6	5
<b>Social Sci. Div.</b>	<b>1,690</b>	<b>1,045</b>	<b>243(16.37)</b>	<b>14.38</b>	<b>23.25</b>		
Sociology	602	381	195	32.39	51.18	3	3
Social Work	292	208	30	10.27	14.42	9	6
Communication	796	456	18	2.26	3.95	10	10
<b>Education Div.</b>	<b>5,304</b>	<b>4,131</b>	<b>226(15.23)</b>	<b>4.26</b>	<b>5.47</b>		
Education	3,541	2,535	186	5.25	7.34	5	7
Physical Education	1,306	1,219	34	2.60	2.79	8	9
Library & Info Sci.	457	377	6	1.31	1.59	11	11
<b>Psychology Div.</b>	<b>1,997</b>	<b>1,779</b>	<b>77(5.19)</b>	<b>3.86</b>	<b>4.33</b>		
Psychology	1,997	1,779	77	3.86	4.33	7	8

\* Rank 1: ranks for the subject divisions; Rank 2: ranks for the subject disciplines based on the proportions of data reuse papers divided by empirical study papers.

The 1,484 data reuse papers together have used 2,990 datasets; that is, each paper has used an average of 2.01 datasets with a standard deviation of 2.02. Observing data usages at the subject division and subject discipline levels, economics research again topped the other subjects (Table 2). Chronological distributions of the dataset usages again showed that there was no obvious increase or decrease of per-paper dataset usages from 2001 to 2015.

In regards to the distributions of data sources and data types, Table 3 shows that data outsourced from government agencies constituted 53.88% of the entire used data, followed by those from academic institutions (18.70%).

**Table 2 The Distribution of Dataset Usages among the Data Reuse Papers, 2001-2015**

Subject Division/ Subject Discipline	No. of reuse papers (% - total reuse papers)	No. of datasets (% - total datasets)	Avg.	St.D.	Rank 1*	Rank 2*	Rank 3*
<b>Total</b>	<b>1,484(100.00)</b>	<b>2,990(100.00)</b>	<b>2.01</b>	<b>2.02</b>			
<b>Economics Div.</b>	<b>597(40.23)</b>	<b>1,504(50.30)</b>	<b>2.52</b>	<b>2.38</b>	<b>1</b>		
Economics	395	1,001	2.53	2.21		1	1
Agricultural Eco.	202	503	2.49	2.68		2	2
<b>Political Sci. Div.</b>	<b>341(26.62)</b>	<b>694(23.21)</b>	<b>2.04</b>	<b>2.11</b>	<b>2</b>		
Politi Sci.	187	389	2.08	2.31		3	3
Publical Admin. & Int'l Affairs	154	305	1.98	1.85		4	5
<b>Social Sci. Div.</b>	<b>243(16.37)</b>	<b>385(12.88)</b>	<b>1.58</b>	<b>1.39</b>	<b>3</b>		
Sociology	195	291	1.49	1.16		5	7
Social Work	30	57	1.90	2.07		8	6
Communication	18	37	2.06	2.07		10	4
<b>Education Div.</b>	<b>226(15.23)</b>	<b>297(9.93)</b>	<b>1.31</b>	<b>1.12</b>	<b>5</b>		
Education	186	243	1.31	1.20		6	10
Physical Education	34	45	1.32	0.68		9	11
Library & Info Sci.	6	9	1.50	0.84		11	7
<b>Psychology Div.</b>	<b>77(5.19)</b>	<b>110(3.68)</b>	<b>1.43</b>	<b>1.13</b>	<b>4</b>		
Psychology	77	110	1.43	1.13		7	9

\* Rank 1: ranks for the subject divisions by the average dataset usage;  
 Rank 2: ranks for the subject disciplines by the total number of dataset;  
 Rank 3: ranks for the subject disciplines by the average dataset usage.

As to data type, Table 4 shows that business data constituted the majority of the used data (54.31%), followed by series surveys (32.31%). The use of one-time data was extremely rare (11.97%).

Based on the analysis, this study concluded that, from 2001 to 2015, slightly lower than one fifth (17.38%) of the social science empirical research in Taiwan was based on the reuse of existing datasets. During the 15 years, both paper production based on data reuse as well as the quantity of datasets used have remained steady. This is possibly the capacity limit of data reuse in generating novel analysis for Taiwan social science research.

Chi-square tests showed that, for both data source and data type of the reused datasets, significant differences existed among the five social sciences subject divisions as well as six major subject disciplines. This means that the data needs and the data reuse behaviors are highly diverse and heterogeneous in different social science subject fields. Economics and political science research constituted the major users of existing data. It was possibly due to the prevalent macro-level research inquiries in those fields, and outsourcing became the only means for data acquisition, particularly for the large-scaled datasets.

**Table 3 The Distribution of Data Sources for the Data Reuse Papers, 2001-2015**

Data Source Div./Discipline	Government Agencies		Academic Institutions		Private Sectors		Individual Researchers		Un-recognized	
	n.	%	n.	%	n.	%	n.	%	n.	%
<b>Total (N=2,990)</b>	<b>1,611</b>	<b>53.88</b>	<b>559</b>	<b>18.70</b>	<b>458</b>	<b>15.32</b>	<b>255</b>	<b>8.53</b>	<b>107</b>	<b>3.58</b>
<b>Eco. Div. (n=1,504)</b>	<b>1,001</b>	<b>66.56</b>	<b>75</b>	<b>4.99</b>	<b>302</b>	<b>20.08</b>	<b>58</b>	<b>3.86</b>	<b>68</b>	<b>4.52</b>
Economics (n=1,001)	666	66.53	51	5.09	224	22.38	24	2.40	36	3.60
Agri. Eco. (n=503)	335	66.60	24	4.77	78	15.51	34	6.76	32	6.36
<b>Poli. Sci. Div. (n=694)</b>	<b>255</b>	<b>36.74</b>	<b>239</b>	<b>34.44</b>	<b>62</b>	<b>8.93</b>	<b>117</b>	<b>16.86</b>	<b>21</b>	<b>3.03</b>
Poli. Sci. (n=389)	119	30.59	132	33.93	41	10.54	87	22.37	10	2.57
Pub. Admin. & Int'l Affairs (n=305)	136	44.59	107	35.08	21	6.89	30	9.84	11	3.61
<b>Social Sci. Div. (n=385)</b>	<b>202</b>	<b>52.47</b>	<b>125</b>	<b>32.47</b>	<b>28</b>	<b>7.27</b>	<b>23</b>	<b>5.97</b>	<b>7</b>	<b>1.82</b>
Sociology (n=291)	138	47.42	114	39.18	14	4.81	19	6.53	6	2.06
Soc. Work (n=57)	51	89.47	3	5.26	1	1.75	1	1.75	1	1.75
Comm. (n=37)	13	35.14	8	21.62	13	35.14	3	8.11	0	0.00
<b>Edu. Div. (n=297)</b>	<b>118</b>	<b>39.73</b>	<b>78</b>	<b>26.26</b>	<b>49</b>	<b>16.50</b>	<b>43</b>	<b>14.48</b>	<b>9</b>	<b>3.03</b>
Education (n=243)	93	38.27	77	31.69	27	11.11	43	17.70	3	1.23
Phy. Edu. (n=45)	21	46.67	0	0.00	18	40.00	0	0.00	6	13.33
Lib & Info (n=9)	4	44.44	1	11.11	4	44.44	0	0.00	0	0.00
<b>Psycho. Div. (n=110)</b>	<b>35</b>	<b>31.82</b>	<b>42</b>	<b>38.18</b>	<b>17</b>	<b>15.45</b>	<b>14</b>	<b>12.73</b>	<b>2</b>	<b>1.82</b>
Psychology (n=110)	35	31.82	42	38.18	17	15.45	14	12.73	2	1.82

**Table 4 The Distribution of Data Types in the Data Reuse Papers, 2001-2015**

Data Type Div./Discipline	Business Data		Series Surveys		One-time Data		Un-recognized	
	n.	%	n.	%	n.	%	n.	%
<b>Total (N=2,990)</b>	<b>1,624</b>	<b>54.31</b>	<b>966</b>	<b>32.31</b>	<b>358</b>	<b>11.97</b>	<b>42</b>	<b>1.40</b>
<b>Eco. Div. (n=1,504)</b>	<b>1,126</b>	<b>74.87</b>	<b>266</b>	<b>17.69</b>	<b>85</b>	<b>5.65</b>	<b>27</b>	<b>1.80</b>
Economics (n=1,001)	739	73.83	197	19.68	42	4.20	23	2.30
Agri. Eco. (n=503)	387	76.94	69	13.72	43	8.55	4	0.80
<b>Poli. Sci. Div. (n=694)</b>	<b>241</b>	<b>34.73</b>	<b>273</b>	<b>39.34</b>	<b>174</b>	<b>25.07</b>	<b>6</b>	<b>0.86</b>
Poli. Sci. (n=389)	95	24.42	174	44.73	115	29.56	5	1.29
Pub. Admin. & Int'l Affairs (n=305)	146	47.87	99	32.46	59	19.34	1	0.33
<b>Social Sci. Div. (n=385)</b>	<b>121</b>	<b>31.43</b>	<b>208</b>	<b>54.03</b>	<b>54</b>	<b>14.03</b>	<b>2</b>	<b>0.52</b>
Sociology (n=291)	74	25.43	172	59.11	43	14.78	2	0.69
Soc. Work (n=57)	29	50.88	20	35.09	8	14.04	0	0.00
Comm. (n=37)	18	48.65	16	43.24	3	8.11	0	0.00
<b>Education Div. (n=297)</b>	<b>100</b>	<b>33.67</b>	<b>161</b>	<b>54.21</b>	<b>32</b>	<b>10.77</b>	<b>4</b>	<b>1.35</b>
Education (n=243)	59	24.28	151	62.14	30	12.35	3	1.23
Phy. Edu. (n=45)	35	77.78	8	17.78	1	2.22	1	2.22
Lib. & Info. (n=9)	6	66.67	2	22.22	1	11.11	0	0.00
<b>Psychology Div. (n=110)</b>	<b>36</b>	<b>32.73</b>	<b>58</b>	<b>52.73</b>	<b>13</b>	<b>11.82</b>	<b>3</b>	<b>2.73</b>
Psychology (n=110)	36	32.73	58	52.73	13	11.82	3	2.73

Also note that data from previous individual research as well as data that were one-time in nature (non-continuous, slice-of-time data) were found to be rarely reused in social sciences. A policy implication from this finding is that the collection development of data repositories for social sciences should focus on acquiring large-scaled surveys as well as business transaction data generated from important governments and private sectors rather than focusing on individual research data that provide very limited reuse opportunities for future researchers.

### **ROMANIZED & TRANSLATED REFERENCE FOR ORIGINAL TEXT**

- 林奇秀、賴璟毅 (2017)。臺灣社會科學學者資料再用行為之研究。圖書資訊學研究，11(2)，95-138。【Lin, Chi-Shiou, & Lai, Ching-Yi (2017). Data reuse behavior among Taiwan social scientists. *Journal of Library and Information Science Research*, 11(2), 95-138. (in Chinese)】
- 科技部人文社會科學研究中心 (2015)。TSSCI收錄期刊名單。檢自<http://www.hss.ntu.edu.tw/model.aspx?no=67>【Research Institute for the Humanities and Social Sciences, Ministry of Science and Technology. (2015). TSSCI shoulu qikan mingdan. Retrieved from <http://www.hss.ntu.edu.tw/model.aspx?no=67> (in Chinese)】
- 陳雪華 (2012)。導論。在陳雪華、陳光華 (編著)，*e-Research* 學術圖書館創新服務 (頁 1-13)。台北市：國立臺灣大學圖書館。【Chen, Hsueh-Hua (2012). Daolun. In Hsueh-Hua Chen & Kuang-Hua Chen (Eds.), *e-Research xueshu tushuguan chuangxin fuwu* (pp. 1-13). Taipei: National Taiwan University Library. (in Chinese)】
- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059-1078. doi:10.1002/asi.22634
- Broom, A., Cheshire, L., & Emmison, M. (2009). Qualitative researchers' understandings of their practice and the implications for data archiving and sharing. *Sociology*, 43(6), 1163-1180. doi:10.1177/0038038509345704
- Chao, T. C. (2012, February). *Exploring the rhythms of scientific data use*. In *Proceedings of the 2012 iConference* (pp. 129-135). New York, NY: ACM. doi:10.1145/2132176.2132193
- Data Curation Center. (n.d.). What is data curation? Retrieved from <http://www.dcc.ac.uk/digital-curation/what-digital-curation>
- Data Information Specialist Committee - UK. (2010). Questions & answers. Retrieved from <http://www.disc-uk.org/qanda.html>
- Fear, K. M. (2013). *Measuring and anticipating the impact of data reuse* (Unpublished doctoral dissertation). Information and Library Studies, University of Michigan, Ann Arbor, MI.
- Henderson, M. E., & Knott, T. L. (2015). Starting a research data management program based in a university library. *Medical Reference Services Quarterly*, 34(1), 47-59. doi:10.1080/02763869.2015.986783
- Jeng, W. (2017). *Qualitative data sharing practices in social sciences* (Unpublished doctoral dissertation). School of Information Sciences, University of Pittsburgh, Pittsburgh, PA.
- Kim, J. (2013). Data sharing and its implications for academic libraries. *New Library World*,

- 114(11/12), 494-506. doi:10.1108/NLW-06-2013-0051
- Konkiel, S. (2013). Tracking citation and altmetrics for research data: Challenges and opportunities. *Bulletin of the American Society for Information Science and Technology*, 39(6), 27-32. doi:10.1002/bult.2013.1720390610
- Kruse, F., & Thestrup, J. B. (2014). Research libraries' new role in research data management, current trends and visions in Denmark. *LIBER Quarterly*, 23(4), 310-335. doi:10.18352/lq.9173
- MacMillan, D. (2014). Data sharing and discovery: What librarians need to know. *Journal of Academic Librarianship*, 40(5), 541-549. doi:10.1016/j.acalib.2014.06.011
- Nasatir, D. (1967). Social science data libraries. *The American Sociologist*, 2(4), 207-212.
- National Research Council. (1999). *A question of balance: Private right and the public interest in scientific and technical databases*. Washington, DC: National Academies Press.
- Piwowar, H. A. (2008). Proposed foundations for evaluation data sharing and reuse in the biomedical literature. *Bulletin of IEEE Technical Committee on Digital Libraries*, 4(2). Retrieved from <http://www.ieee-tcdl.org/Bulletin/v4n2/piwowar/piwowar.html>
- Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, e175. doi:10.7717/peerj.175
- Piwowar, H. A., Carlson, J. D., & Vision, T. J. (2011). Beginning to track 1000 datasets from public repositories into the published literature. *Proceedings of the American Society for Information Science and Technology*, 48(1), 1-4. doi:10.1002/meet.2011.14504801337
- Robinson-García, N., Jiménez-Contreras, E., & Torres-Salinas, D. (2016). Analyzing data citation practices using the data citation index. *Journal of the Association for Information Science and Technology*, 67(12), 2964-2975. doi:10.1002/asi.23529
- Sechrest, L., & Sidani, S. (1995). Quantitative and qualitative methods: Is there an alternative? *Evaluation and Program Planning*, 18(1), 77-87. doi:10.1016/0149-7189(94)00051-X
- Van den Berg, H. (2005). Reanalyzing qualitative interviews from different angles: The risk of decontextualization and other problems of sharing qualitative data. *Forum: Qualitative Social Research*, 6(1). doi:10.17169/fqs-6.1.499
- Xia, J., & Liu, Y. (2013). Usage patterns of open genomic data. *College & Research Library*, 74(2), 195-207. doi:10.5860/crl-324
- Zhang, J. (2011). *Data use and access behavior eScience — Exploring data practices in the new data-intensive science paradigm* (Unpublished doctoral dissertation). Department of Information Science, Drexel University, Philadelphia, PA.
- Zhang, J., Vogeley, M. S., & Chen, C. (2011). Scientometrics of big science: A case study of research in the Sloan Digital Sky Survey. *Scientometrics*, 86(1), 1-14. doi:10.1007/s11192-010-0318-1

