

臺灣社會科學學者資料再用行爲 之研究

Data Reuse Behavior among Taiwan Social Scientists

林奇秀

Chi-Shiou Lin

國立臺灣大學圖書資訊學系暨研究所副教授
Associate Professor
Department of Library and Information Science,
National Taiwan University

賴璟毅

Ching-Yi Lai

國立臺灣大學圖書資訊學研究所研究生
Graduate Student
Graduate Institute of Library and Information Science,
National Taiwan University

【摘要 Abstract】

近年來學術資料分享呼籲漸盛，學者的資料再用行爲也開始受到關注，但相對於自然與應用科學，社會科學的資料再用研究相對較少。本文以深度訪談來瞭解臺灣社會科學學者的資料再用行爲，共訪談 14 位具有量化數據再用經驗的社會、政治、教育、經濟與心理學者。研究發現，在再用動機方面，學者取用既有資

料，理由可能是因為無法自行蒐集大規模或長期性資料、取用來自權威單位的資料比較具有公信力、取用既有資料可免除研究倫理審查的煩擾、可探索潛在研究題目、能拓展既存的研究議題、學科領域文化的鼓勵或阻撓等。在獲知管道方面，學者可能透過五種管道來掌握資料的消息，包含學術文獻、同儕與指導教授、政府與學術機構網站、學會或調查機構的推廣活動、紙本統計資料等。在資料評估方面，學者常用的評估策略包含評估問卷工具品質、評估樣本品質、評估資料蒐集過程、評估資料新穎度、評估資料易得性、以及評估分析結果的發表潛力。在資料處理上，學者會先對資料集進行描述統計與信效度檢驗，確認資料基本品質後，再進行基本數據校正處理，有時還必須合併多組資料集並尋求或補足資料集缺乏的變數。

This article describes the findings from a qualitative study on social scientists' data reuse behavior in Taiwan. In recent years, data sharing has been a salient topic in the scholarly communities. Empirical studies on how scientists use existing data have also emerged. However, data reuse in social sciences is less studied than in natural and applied sciences. This study focused on the reuse of existing quantitative data by Taiwan's social scientists. Semi-structured, in-depth interview was used to understand the experiences of 14 researchers from sociology, political science, education, economics, and psychology. The results show that, in regards to motivations, participants re-used existing data for the following reasons: unable to collect large-scaled or long-term data, higher credibility of data released by authoritative sources, free from harassment of IRB reviews, exploring potential research questions, carrying on existing research directions, and the encouragement or discouragement of the disciplinary cultures. Participants had relied on five different channels to find existing data, i.e., research literatures, peers and advisors, government agencies or academic institutions' websites, promotional activities of scholarly associations and survey institutes, and statistical publications. Participants also reported six evaluation strategies prior to the actual use of the located data, including the evaluation of survey instrument quality, sample quality, data collection procedures,

timeliness of data, availability of data, and the potential value for publication of the re-use analysis. Finally, in data processing, participants first inspected the descriptive statistics and the validity and reliability of the dataset and then proceeded to correct, amend, or convert the data. They might also need to combine multiple datasets and fill in further needed variables from external sources or by using various techniques.

【關鍵字 Keywords】

資料再用、資料取用、量化資料、社會科學

Data reuse; Data access; Quantitative data; Social sciences

壹、前言

學術界的資料分享（data sharing）係指學者釋出一手資料，透過公開的典藏機制供他人再利用（Borgman, 2012）。資料分享運動旨在促成資料再用，倡議者認為資料的公開分享與再用有三大好處：首先，資料再用可節省研究成本：除可避免學者重複蒐集性質相近的研究資料（Konkiel, 2013; Kvalheim, Kvamme, & Norwegian Social Science Data Services, 2014），學者亦可對既有的資料進行新的分析，發揮最大的研究效益（Fienberg, Martin, & Straf, 1985; Butler, 2006; Whitlock, 2011; Zenk-Moltgen & Lepthien, 2014）；其二，資料再用促成研究分析的公開透明：學術社群得以透過原始資料來檢驗研究結果，避免資料曲解（Fienberg, Martin, & Straf, 1985; Butler, 2006; Whitlock, 2011; Zenk-Moltgen & Lepthien, 2014）；其三，資料再用可增進公共利益：資料可供學術界以外的社會大眾使用，例如作為學生學習數據分析之素材，或提供行政單位作為施政參考等（Fienberg et al., 1985; Whitlock, 2011）。

雖然資料分享是近年才扶搖直上的議題，但在社會科學領域中，資料分享與再用行為其實已經存在五十年以上了。歐美國家在 1960 年代即意識到社會科學數據資料保存與再用的需求，也出現各種資料典藏機構，當時各國合計就有廿餘所，知名者如美國密西根大學發起的跨校資料典藏服務 Inter-university Consortium for Political Research（ICPR）、加州大學柏克萊分校之 International Data Library and

Reference Service、美國聯邦政府的普查局與勞動統計局、私人民意調查機構 Roper Public Opinion Research Center 等；歐洲則有德國的 DATUM 與 Zentralarchiv für Empirische Sozialforschung（社會科學的資料中心）、英國的 Social and Economic Archive Committee、荷蘭的 Steinmetz Institute 等（Nasatir, 1967）。

其後，聯合國於 1971 年提出 UNISIST Model。UNISIST 全名為 United Nations International Scientific Information System，在這個科學資訊的生產與交換體系中，資料服務被視為學術傳播體系中必要的一環（Fjordback Søndergaard, Andersen, & Hjørland, 2003），然而，在 70 年代以降的學術傳播研究，關注焦點大幅集中在研究成果的傳播與交換，如科學論文與報告，較少關注資料的分享與取用。實務上，無論在學術圖書館或資訊檢索的發展也多以論文成果為主軸，甚少關注原始資料的傳播與交換。直到 2000 年中期，因資料皮用意識抬頭，資料再用與相關服務才開始真正受到矚目（Xia & Wang, 2014）。

受到 E-Science 與資料皮用意識的影響，學術資料的再用研究也漸受關注，但既有的實徵研究多以自然理工學科為對象，尤其是資料分享與再用較盛的基因領域（Piwowar, 2008; Piwowar & Vision, 2013; Xia & Liu, 2013）與天文學（Zhang, 2011, Zhang, Vogeley, & Chen, 2011），僅有 Fear（2013）曾以知名社會科學資料服務機構 Interuniversity Consortium for Political and Social Research（ICPSR，前身即為 ICPR）為對象，分析該機構提供的資料被學者再用的狀況；Curty（2015）與 Niu 與 Hedstrom（2009）則對社會科學學者進行調查，瞭解其常用資料型態、資料取用來源與偏好。在研究方法上，前述文獻均以量化分析來探討再用論文與被用資料的特徵或資料再用的傾向，較少關注學者再用的行為本身，例如學者如何發現、評估、與使用他人提供的資料。簡言之，我們目前對學者的資料再用行為瞭解甚少，這個問題值得深究。

社會科學研究所使用的資料形式非常多元，根據 Data Information Specialist Committee UK（2010, as cited in Kruse & Thestrup, 2014）的例舉，凡數字、統計、表格、圖像、影片、地理空間資訊等均可作為資料；而且，無論是數字、文字、符號、甚或是實物等任一形式，只要其能描述或反映學者關懷的現象、事件、思想、情況、或處境，就是研究資料（National Research Council, 1999）。不同的資料有不同的使用目的與再用條件，很難一概而論，本文在探究社會科學學者的再用行為時，故以量化資料為限，一來因該類資料為社會科學資料再用的大宗

(Curry, 2015)；二來，質化資料因與原始研究命題關係緊密，關於質化資料分享的效益與再用效力，各方見解尚不一 (Mauthner & Parry, 2009; Van den Berg, 2005)，本文故以再用已具制度規模的量化資料為探討重點。

具體而言，本文的研究關懷是社會科學學者「為何」與「如何」使用他者提供的量化資料，包含政府或科研機構從事的調查資料、公私單位提供的業務統計、前人研究釋出的原始量化數據等。本文的具體研究問題包含：(1) 學者基於哪些動機或理由來再用資料？(2) 有哪些獲知資料的管道？(3) 如何評估資料是否可用？(4) 進行分析前，又必須對資料進行哪些處置？透過對 14 位社會科學學者的半結構訪談，本文將依序說明對上述四個問題的研究發現，期能對我國推動資料分享與再用有所裨益。

貳、文獻回顧

本文共探索四個研究問題，然在回顧文獻時，發現文獻中雖有針對資料再用動機、獲知管道、與資料評估等相關實徵研究，但尚未見有研究探討再用分析前的資料處理行為。本節故呼應前三個研究問題，綜述前人發現：

一、資料再用動機

在自然與應用科學方面，文獻指出既有資料的再用動機至少包含三點：(1) 利用既有資料來補充或強化現行研究的資料基礎，減少自行收集資料的成本 (Wynholds, Fearson, Borgman, & Traweek, 2011; Wallis, Rolando, & Borgman, 2013)；(2) 用來測試研究模型或工具：以地震科學為例，學者在發展探測儀器或預測模型時，為檢驗模型或方程式的效度，會利用既有資料來進行測試 (Faniel & Jacobsen, 2010b; Wallis, et al., 2013)；(3) 補充背景知識，熟悉研究主題：以天文學及應用科學為例，學者可能會先試行分析前人所產出的研究資料，藉以熟悉其預定研究的主題或場域 (Wynholds, et al., 2011; Wallis, et al., 2013)。在社會科學方面，Yoon (2015) 訪談社工與公共衛生學者，發現使用既有資料的理由包含：(1) 既有資料的內容本身足以回答研究問題，毋須耗費額外的成本再蒐集資料；(2) 某些既有資料樣本數量龐大，無法靠個人研究蒐集。

二、獲知管道

學者獲知資料存在的途徑包含人際管道、數據庫或期刊論文、資料說明書。首先，在人際管道方面，指導教授、同儕、及專業機構人員是三種主要人際途徑。Kriesberg、Frank、Faniel 與 Yakel（2013）發現指導教授是指引年輕學者資料所在，並指導資料使用細節與技術的重要角色。Daniels（2014）發現同儕也是資料訊息的主要來源，而當同儕即是資料生產人時，還能提供資料細節與使用諮詢，例如：地震學者會向同儕詢問資料蒐集儀器、型號、材質及資料信效度等（Faniel & Jacobson, 2010b），社會科學學者則會徵詢質化資料產生的情境脈絡（Yoon, 2014）。專業機構人員如博物館員或資料圖書館的館員等，除可提供一手資料所在地、資料細節消息等，也可能會協助資料整理、轉檔、萃取等資料整理分析工作（Read, 2007; Faniel, Kansa, Kansa, Barrera-Gomez, & Yakel, 2013; Daniels, 2014）。

其次，對某些資料再用頻率較高的學科來說，領域內的知名數據庫也是重要的消息來源，例如基因研究中的 GenBank（Wallis et al., 2013），或天文領域的 Sloan Digital Sky Survey（SDSS）（Sands, Borgman, Wynholds, & Traweck, 2012）。此外，學者也會透過期刊論文來發現已經存在的資料，例如，原始論文對資料集的指引與說明，可讓後續再用該批資料的學者理解資料的蒐集方式、內容與限制（Faniel & Jacobsen, 2010b），而資料如經再用，再用論文對資料的分析與評論，也可以讓後續學者評估該筆資料是否仍有分析價值（Faniel, Kriesberg, & Yakel, 2012; Faniel, Barrera-Gomez, Kriesberg, & Yakel, 2013）。

另一個獲知管道是資料說明書（documentation），這類文件是資料生產者對資料蒐集、測試、分析、與處理過程的紀錄（Faniel & Jacobsen, 2010b），其資訊豐富程度，通常比數據庫的書目紀錄或原始論文的說明要來得更好，學者們在取用資料時，多會參考資料說明書（Carlson & Anderson, 2007; Niu, 2009; Niu & Hedstrom, 2009; Faniel & Jacobsen, 2010b; Faniel & Barrera-Gomez et al., 2013; Stvilia et al., 2015）。

三、資料評估

既有文獻較少呈現學者對既存資料的評估與取用決策行為，但有數篇實徵研究在探討哪些因素會影響學者是否再用既存資料。綜合不同研究的發現，可發現資料的正確性（accuracy）或可信性（trustworthiness 或 credibility）、可及性（accessibility）、可用性

(usefulness)、與相關性(relevance)等是比較重要的因素，以下簡述之：

(一) 正確性／可信性

不同的研究在探討資料再用的影響因素時，對影響因子的概念建構或操作定義多有差異，但綜觀不同研究的發現，可發現資料的正確性或可信性，可能是影響再用最重要的因素。Huang、Stvilia、Jorgensen 與 Bass (2012) 針對 158 位基因學者進行調查，透過因素分析，將資料的「正確度」、「無偏誤」(unbiased)、「可信度」(believability)、「可追蹤度」(traceability) 等測量題項，均歸納在資料正確性這個構面下；Stvilia 等人 (2015) 針對 172 位凝聚體物理學者的調查研究，也以因素分析將資料的「正確性」、「完整性」、「一致性」(consistency)、「精確性」(precision)、「信度」、「效度」等歸納在正確性之下。在這兩個研究中，正確性對資料品質判斷的解釋力均是最高的，且 Stvilia 等人發現，在 13 種資料品質判斷的指標中，正確性位居首位。

可信性的概念與正確性高度相關，係指學者認為資料值得相信的程度。Stvilia 等人 (2015) 發現，在 13 種影響品質判斷的指標中，可靠性(reliability)與可驗證性(verifiability)排序分列第二與第三名，而這兩個指標都與資料的可信性有關。Faniel 等人 (in press) 則以資料是「可信的」(credible)、「可靠的」(reliable)及「是被客觀蒐集出來的」等三個問項來測量資料可信度，結果顯示可信度與使用滿意度呈顯著正相關。

質化研究也常提及學者對資料可信性的判斷行為。一般而言，資料的背景資訊愈豐富，學者對資料的信任程度就愈高，特別關於數據產製之研究方法設計(Niu & Hedstrom, 2009; Faniel, et al., 2012; Murrilo, 2014)及儀器工具細節(Mayernik, Wallis, Pepe, & Borgman, 2008; Faniel & Jacobsen, 2010a, b)。此外，學者也會依據資料來源的聲望或權威度來判斷資料是否可信(Zimmerman, 2008; Gamble & Goble, 2011; Wynholds et al., 2011; Huang et al., 2012; Faniel & Kansa et al., 2013; Curty & Qin, 2014)。有時候，領域內的流派差異，對資料可信性的判斷也會造成影響，例如考古學者便會觀察資料作者及其所屬機構的學派風格，來決定該批資料是否值得信賴、適合己用(Faniel & Kansa et al., 2013)。

(二) 可及性

可及性是指能尋得並獲取資料的程度。Stvilia 等人 (2015) 對 13

個資料品質指標進行因素分析時，發現可及性這個單一指標自己形成一個獨立構面；而 Huang 等人（2012）則透過因素分析，發現可及性構面下，包含了「可取得」、「可信的」、「有適量資訊」（appropriate amount of information）等指標；Faniel 等人（in press）對 237 位社會科學學者的調查研究，則以資料是「容易檢索的」、「容易取用的」、「容易獲得的」來測量可及性。就研究結果來說，Huang 等人與 Stivilla 等人均發現，在各種影響資料品質判斷的構面中，可及性均位居第二，僅次於正確性構面，可見其重要性；而 Faniel 等人透過多元迴歸分析，也發現可及性對資料的使用滿意度具有解釋力。

此外，Curry（2015）對 558 位美國社會科學學者進行調查，試圖瞭解各類變數對資料再用意圖的影響。研究發現「促進情境」（facilitating condition）對資料再用有正面影響，而細觀所謂的促成情境，多數與資料可及性有關，例如「資料說明書的可得性」、「數據庫的可及性」、「是否可聯絡到資料原作者」、「其它使用支援可得性」等。由此可知，可及性的確是相當重要的影響因素。

（三）可用性

可用性係指學者對資料有用程度（Huang et al., 2012）或好用程度（ease-of-use）（Niu & Hedstrom, 2009）的判斷，這牽涉到資料內容是否好理解、資料類型或資料格式是否容易使用等。Huang 等人（2012）透過因素分析，將資料的「可詮釋性」、「易理解性」、「容易操作」、「是否經過加值處理」等歸納為可用性；Faniel 等人（in press）以資料是否「容易操作」、「容易整合」來測量可用性；Curry 與 Qin（2014）及 Curry（2015）對社會科學學者的質、量化研究中，也曾以「資料說明書的品質」、「資料的適用性（fitness）」等，測量學者對資料感知的可再用性（perceived reusability）。

在研究結果上，Huang 等人（2012）發現可用性在各種影響資料品質判斷的因素中，重要性排序第三；Faniel 等人（in press）發現可用性對資料再用滿意度具有正面解釋力。Curry 與 Qin（2014）對 13 位社會科學研究者的深度訪談顯示，學者所感知的可再用性，會對資料再用意圖產生正面影響，但 Curry（2015）後續以結構方程式對 558 位美國社會科學學者的調查結果進行分析，卻顯示該因素不會影響再用的意圖，他推測這個結果，可能是因為測量題項的聚合效度（convergent

validity) 不佳而導致的。

(四) 相關性

相關性大致上是指資料與研究主題或研究需求的契合程度，相對於前述幾個因素，相關性的概念比較模糊，操作測量方式也高度分歧，與其他因素重疊的程度頗高。Faniel 等人 (in press) 以學者是否感覺資料「有用」或「適合」(appropriateness) 來測量相關程度；但 Huang 等人 (2012) 基於因素分析，將資料是否具有「代表性」、「為最新版本」、「產製者聲望」、與「被加值處理過」等，全部歸納為相關性。上述研究操作方式顯示：相關性作為一個影響因子，無論在概念清晰度或分析解釋力，都遠不如前述因素，例如 Stivilia 等人 (2015) 用來測量資料品質的 13 個指標中，原本有相關性這個項目，但在進行因素分析後，相關性就被吸收到資料的「資訊豐富度」(informativeness) 底下，不作為一個單獨解釋的因子。

作為解釋因子，相關性的重要性雖不如前述各種因素 (Huang et al., 2012; Stivilia et al., 2015; Faniel et al., in press)，但學者在取用資料時，勢必會挑選主題合適或感覺有用的資料，因此，相關度幾乎是選擇資料的前提，因此也仍是影響再用與否的因子。

上述研究成果能解釋學者在選擇資料時有哪些品質訴求，但在行動層次上，則較少呈現學者如何觀察與評估資料，以判斷資料品質是否滿足訴求。本文故以質化訪談來瞭解學者的資料評估行為，剖析其關懷面向與重點，期能與文獻呼應。

參、研究方法

本文以半結構深度訪談來瞭解社會科學學者如何取用外部的量化資料來從事具原創性的研究。在受訪者招募上採立意抽樣，以 2015 年的 TSSCI 收錄期刊為基礎，尋求在 2011-2015 年間，曾於社會、政治、教育、經濟與心理這五個學門期刊中發表再用論文的學者，並由近期發表的論文開始，根據論文所提供的電子郵件資訊，邀請符合條件的論文第一作者受訪。

本研究原訂每個領域至少訪談三位學者，共寄發出 41 封邀訪信 (社會 8 封、政治 8 封、教育 3 封、經濟 17 封、心理 5 封)，最後，前四個領域均成功訪談到三位學者，唯心理領域再用論文本來就不多 (2011-2015 年間全部僅 17 篇再用論文)，且主題幾乎都與教育領域

重複，最後僅有兩位學者同意受訪，且其中一位的主題仍與教育相近。

全部 14 位受訪者皆曾用過來自政府、學術或民間機構的系列調查資料（亦即有分多年重複執行的大規模調查）。曾使用政府機關或民間機構的日常業務統計者，僅有一位社會學者與三位經濟學者，但受訪的經濟學者表示，業務統計是經濟研究的重要數據。僅有兩位學者曾使用來自前人個別研究的一次性資料，且這類資料也不是其再用經驗的大宗。個別研究產出的資料再用，在社會科學是比較少見的（參見表 1）。

表 1.
受訪者基本資料

受訪者代號	受訪者領域	研究年資	使用資料類型		
			系列調查	業務資料	個別研究
社 A	社會學	10 餘年	✓		
社 B	社會學	7 年	✓		
社 C	社會學	20 餘年	✓	✓	
政 A	政治學	10 餘年	✓		
政 B	政治學	15 年	✓		
政 C	政治學	10 餘年	✓		
教 A	教育學	10 餘年	✓		
教 B	教育學	5 年	✓		
教 C	教育學	4 年	✓		
經 A	經濟學	20 年	✓	✓	
經 B	經濟學	10 餘年	✓	✓	✓
經 C	經濟學	18 年	✓	✓	
心 A	心理學	3 年	✓		✓
心 B	心理學	10 餘年	✓		

註：「系列調查」是指由政府、學術或民間機構分多年期執行的調查資料；「業務資料」指政府機關或民間機構日常行政業務的統計數據；「個別研究」則是指從前人執行的一次性個別研究中所產出的數據。

訪談內容聚焦於受訪者的資料再用動機、如何獲得並評估資料、在分析前對資料進行哪些前置處理等。研究者在訪談中，會針對受訪者近期發表的特定再用論文及其一般資料再用行為進行詢問，並

會請其說明或評論其所屬的學術領域資料再用狀況（訪談大綱如表 2）。訪談均為一對一的面對面訪談，每場訪談介於 60-120 分鐘，全程錄音並抄錄成逐字稿，後續在資料分析時，則參酌 Strauss & Corbin（1998/2001）的質化研究編碼方法，先以開放編碼辨識各種資料再用的動機、理由、獲知管道、評估面向、處理程序等，後續再遵循研究問題的引導，以軸心編碼將前述開放編碼的成果進行歸納、聯結、與統合，形成研究發現。

表 2
訪談大綱

核心概念	訪談問題
再用動機 或理由	<ul style="list-style-type: none">● 請問您在近期這篇論文中使用這些資料的理由是什麼？● 就您個人經驗而言，使用既有資料進行研究有什麼好處？● 就您所屬領域而言，使用資料進行二次分析是否普遍？為什麼？
獲知管道	<ul style="list-style-type: none">● 一般而言，您是怎麼知道資料的存在的？在您近期發表的這篇論文中，您又是透過什麼管道得知的呢？● 在不同的資料獲取管道中，何者對您而言比較重要？為什麼？
資料評估	<ul style="list-style-type: none">● 您會從哪些面向去評估資料是否可用？
資料處理	<ul style="list-style-type: none">● 您在實際分析資料之前，必須對資料進行哪些處理工作？

肆、研究發現

一、量化資料再用的動機或理由

根據本文的訪談顯示，社會科學學者使用既有的量化資料的動機，包含學者無法自行收集所需資料、來自權威單位的外部資料較具公信力、免除倫理審查的限制、能探索潛在研究題目、拓展既存的研究議題，以及學科領域文化的鼓勵或限制等，以下分別說明：

（一）學者無法自行蒐集所需資料

1. 研究需要宏觀、大量且具代表性的樣本

對於多數受訪者（社 A、社 B、社 C、政 A、政 C、經 A、經 B、

經 C、教 A、教 C、心 A、心 B) 而言，在進行研究時，需要大量且有代表性樣本，或研究範圍為整個國家層級，甚至是跨國型的資料才能回應其研究問題。這些研究資料無法靠單一學者或研究團隊蒐集，僅能透過政府或學術機構的力量才有辦法蒐集，因此學者必須仰賴政府與學術機構蒐集的資料。例如，長年使用跨國樣本的社 C 便表示：「像世界價值觀的資料檔，資料內容跨好幾個不同國家，一般人是沒有辦法蒐集這麼多的。」而經 B 所使用的總體經濟數據也是國家級的資料，僅能從政府機關獲得，他指出「經濟學常用數據如 GDP、失業率啊、通貨膨脹率，那都是國家發布的總體資料。」

另一方面，因為大樣本的蒐集需要鉅額研究經費，也不是單一研究者可以負擔的，社 B、社 C、政 B、政 C 與經 B 等都指出，個人之力幾乎無法蒐集到足量的資料規模：

委託調查機關或各地的調查中心幫我們調查，它一份完整問卷大概要上千塊。你一做就要三、四千份……我們經濟主要是做量化分析，樣本數五六百算是少的，大部份都要一、兩千份樣本（經 B）。

在教育研究方面，許多教育研究所需要的數據必須仰賴各層級學校配合才能完成蒐集，學校比較可能會配合政府主導的調查計畫，所獲得樣本數據會比個人去蒐集要來得更具較代表性：

如果是政府的案子的話，其實政府發文到教育局，教育局就會分配那些學校，讓你做你的施測對象，那你的施測樣本就會更廣（教 C）。

2. 研究需要長期性資料

對於半數受訪者（政 A、政 B、政 C、心 B、經 A、經 C 與教 C）而言，他們需要的研究數據是來自長期主題調查或長期追蹤調查。前者是指用同一問卷題組，逐年進行抽樣調查所獲資料，例如政治學中的選舉研究就高度仰賴這類調查，以政 A 研究的政治信任感為例，因過去數十年來台灣社會與政治變動劇烈，因此在分析社會背景及政治信任感對投票行為的影響時，就需要從歷年的長期主題調查中抽取相關數據出來分析。長期調查對選舉研究這個領域非常重要，因此該領域很早就建立起分工合作、共享調查成果的慣習：

我未來四年每年都會蒐集資料，我自己要來推動，今年、明年、後年都做一次電訪。[但]如果往前看的話，過去年份我不可能自己執行電訪了嘛，[所以取用別人收集的既有資料是必須的]（政 B）。

長期追蹤調查則是針對同一群樣本，進行歷時性的定期調查，長期追蹤除需要耗時進行調查，調查樣本還必須固定，個別研究團隊很難追蹤樣本動向，且容易流失樣本，因此多須仰賴政府或大型機構的力量來執行。受訪者中，研究老人學的心 B、研究學業與成就表現關連的經 A 與教 C、以及研究企業發展的經 C 等，均使用過這類資料。使用長期追蹤資料的好處是能控制與樣本特質相關的變數，如此才能有效檢驗自變數與依變數之間的關係：

你做老化研究時，你要有長期追蹤的樣本，也要有不同年齡層的樣本，你才可以區隔開說，這個差異是不同世代，或是真的隨著年齡過去就會有的變化（心 B）。

（二）來自權威單位的資料較具公信力

多位受訪者（社 A、社 C、政 A、政 B、政 C、經 B 與經 C）也指出，資料如來自政府或具名氣的學術機構，因為有經費與制度支持，調查品質會比個人調查好。以選舉研究為例，該領域公認政治大學選舉研究中心的調查資料品質好，因為是透過面訪所得：

面訪蒐集的資料，品質方面一定會比電訪好，因為面訪是面對面到你家去訪問，我們認為那樣子蒐集的資料會比較具可信度。電訪會遇到很多問題，例如被掛電話，或是抽樣的問題等等（政 B）。

使用這類權威性資料的另一個好處是，投稿時，審稿者也比較不會質疑這些資料的信效度：

如果你論文是用 TSCS（註：臺灣社會變遷基本調查）的資料，事實上，你都不用太多的說明，在學界它已經有公認的信度跟效度，這是為什麼要選它的一個很重要的原因（社 A）。

(三) 可免除研究倫理審查的限制

教 C 另外提到，近年來研究倫理審查愈加嚴格，研究者進行調查須獲得受測者的調查同意書，在大規模研究的情境下，這對個別研究者的負擔太大，如果既有資料中有可以使用的數據，就可以免於研究倫理審查的煩擾：

只要是調查國中或高中以下的學生，我們都要進行研究倫理的審查，這對研究者來講是一件花費心力的事，因為你要取得每一個學生的同意書……如果你用蒐集好的資料庫的話，就不會遭遇這樣的問題（教 C）。

(四) 能探索潛在研究題目

教 A 與教 B 另指出，自行設計問卷去蒐集資料，回收後可能發現資料分析結果與預期不同，或分析後才覺得結果不具發表價值等問題，但調查既已執行完畢，也無法挽回或補救。若是利用現成的既有資料進行分析，就可以避免這類問題。以教育領域而言，教 B 認為政府與學術機構調查的既有資料，就已經包含有相當豐富的變數，足以讓研究者從中觀察並產生各種假設，從事探索性分析：

比如說想測量學生的心理壓力，用這幾個向度就可以代表學生的壓力……[而如果某些] 壓力測量的項目跟我的研究方向是沒有關係的時候，我也可以不使用它，[我] 可以從我的理論設計去作取捨（教 B）。

(五) 能拓展既存的研究議題

本研究受訪者中，曾使用前人在個別研究所產出數據來進行再用分析者較少，僅經 B 與心 A 兩位。經 A 指出，經濟學研究多半是先有問題意識，然後再去尋找合適的外部資料，例如臺灣社會變遷調查資料等，但他也曾有一次取用前人個別研究資料的經驗，該資料來自一個大型研究計畫，其指導教授是該筆資料的問卷設計者，經 A 雖未參與問卷原始設計與分析，但後續則與其教授一同對該筆資料進行衍生議題分析。心 B 則剛好相反，資料來源是其指導學生，因其學生對發展測驗量表有高度興趣，因此他們曾合作發展量表並收集數據，後續遂運用該批數據繼續進行數個不同面向的分析與檢測，也均有獲得研究成果。

（六）學科領域文化的鼓勵或限制

受訪者指出資料再用的程度，與學科領域的風氣有關。政 C 指出，選舉研究領域中資料再用論文很多，這與學科發展歷史有關，當政治學開始朝向量化研究發展，學者就開始體悟到合作調查、數據共享的重要性，因此逐漸形成合作建置大型資料庫的風氣：

大型資料庫出現，是希望整合學界資源……既然我們是研究相似的主題，我們就用同一筆資源，我們共同去研擬開發需要的資料，慢慢建立統合的資料庫……像 TEDS（註：臺灣選舉與民主化調查資料庫）的精神，就是大家統合起來一起做（政 C）。

然而，學科領域文化也可能會限制資料再用的發展，例如教 B 覺得資料再用在教育領域並不興盛，是因為教育領域內負面的官學關係所造成的。具體來說，他認為政府官員在教育政策制訂過程中，並不真正重視科學分析的結果，學者的分析成果常淪為政策背書的工具，導致這個領域中，學者對數據的建立與再用都會感到懷疑與疏離：

我舉一個比較簡單的例子，你如果去找一個高中職社區化的議題，你會發現有兩派的資料，他們分別呈現支持跟反對的立場……基本上他們（註：政府官員）不管研究這塊，反正這就是既定政策，就這樣走下去（教 B）。

在心理學部分，資料再用比例高低，則與治學典範（paradigm）有關。心 A 指出，台灣心理學界中很少見資料再用，因為台灣的心理學者比較偏好用既有理論去詮釋現象，因此也傾向自行發展問卷來收集數據；相對而言，他自身因受到特定心理學派的影響，比較強調對實徵現象的原始觀察，不受既有理論的宰制。在學域中不同典範的彰顯程度，就會明顯影響資料再用行為的實踐程度。

三、獲知資料的管道

在獲知管道的經驗方面，歸納出學術文獻、同儕與指導教授、政府與學術機構網站、學會與調查機構，與紙本統計資料等五種管道，以下簡述之：

(一) 學術文獻

有 12 位受訪者（社 B、社 C、政 A、政 B、政 C、經 A、經 B、經 C、心 A、教 A、教 B 與教 C）曾透過學術文獻來得知特定資料的存在，因為再用論文通常會交代其使用的數據來源及特徵，因此，學術論文本身即是一種重要的資訊來源。

公開資料庫是非常興盛的。有些學術文獻你在接觸過程中，會瞭解到哪些資料庫是可以公開使用的（政 A）。

(二) 同儕與指導教授

半數受訪者（社 A、政 C、教 A、教 B、經 B、經 C 與心 A）曾透過指導教授獲得資料，例如教授授課時曾用的資料、在個別研究或參與大型調查計畫所得資料等。而他們自己成為老師之後，也常在課堂中將資料分享給學生，或告知學生重要資料管道，例如中研院的學術調查研究資料庫（Survey Research Data Archive, SRDA）（社 A）。

有 6 位受訪者（社 B、政 A、政 B、教 A、教 B 與經 B）曾透過學術同儕的交流如讀書會、合作研究，或社交聊天場合得知資料相關資訊：

像我們在用到大量的國民所得的資料時，不是很清楚在哪裡，但我們可以問同事說在主計處哪裡可以找到（經 B）。

(三) 政府與學術機構網站

社 C 與心 A 在尋找政府釋出的業務統計資料時，會從資料主題來推敲業務的權責單位，再至該單位尋找可用資料：

你的題目本身就是在那個範圍（註：國家衛生），你當然去找世界衛生組織……聯合國可能會有一些，世界銀行可能也會有一些，但是最完整的還是應該到世界衛生組織去找（社 C）。

多位受訪者（社 A、教 A、教 C、經 A、經 B、經 C 與心 A）也會透過瀏覽政府或大型知名機構的網站來發掘潛在資料，特別是當已知有特定資料庫所在，延伸瀏覽該機構網站，或許可能會獲得其他具有利用價值的資料：

以國外網站來講，國際公民素養那個，他一定是跟很多的資料庫擺在一起的，那我就一個一個去點，看我們臺灣有沒有做相關的東西或相關的資料庫（教 C）。

（四）學會或調查機構的推廣活動

四位受訪者（心 A、社 B、經 B 與經 C）曾透過中研院或其他學術機構的資料推廣工作坊或研討會等，得知資料數據的存在。這些資料推廣活動會介紹的資料性質、蒐集過程、甚或提示該資料庫可支援哪類研究，好的說明能讓受訪者會對數據庫產生信任感：

我參加過幾次工作坊，那個工作坊會讓我們知道說，那個資料從設計到調查有多少人去參與、他們怎麼建置那個資料庫，從一開始有多少人參與，然後怎麼樣設計問卷，設計出來之後怎麼樣去試問、前測等……我覺得這整個過程會讓我相信這個資料庫是很嚴謹在執行（經 C）。

專業學會與機構寄發的電子報也常報導資料訊息，對受訪者是重要的資料消息來源，例如經 A 說：「我們經濟學界有一個年會，它會 EMAIL 通知我們，說現在有什麼新的資料庫，會介紹給大家使用。」社 C、政 B、教 B 與教 C 也表示，中研院或大型調查中心主動寄發的數據庫介紹、資料光碟片等，是他們掌握資料的管道之一。

（五）紙本統計資料

傳統的圖書館紙本統計資料館藏，也仍是學者查詢數據資料的基本管道。經 C 就提到他在蒐集家庭收支數據時，曾先逐篇翻閱圖書館的相關紙本統計報告，確認官方統計調查項目的詳細程度，評估研究分析的可用性後，再想辦法去索求資料。

四、資料評估

受訪者在評估是否取用既有資料時，會評估其問卷工具品質、樣本品質、資料蒐集過程、資料新穎度、易得性、以及分析結果是否具發表潛力價值等，以下分述之。

（一）評估問卷工具品質

作為研究工具，問卷設計的品質與數據品質息息相關。綜合訪談

資料，受訪者在評估工具品質時，會從以下幾個角度來觀察：

1. 問卷能否回應再用者所關心的問題

所有受訪者皆表示，他們從事資料再用時，首要評估的就是既有資料主題內容能否回應當下所關切的議題，如果資料問項或資料主題不完全吻合再用分析的研究設想時，受訪者也可能會調整自己的分析概念，例如：

我們一開始想要做的是，當他有過度教育的時候，他跟他工作的配合度的關係，就是學用有沒有相符……但是那個調查，並沒有調查工作配合度，所以後來我們才會改成去看他求職過程的時間長短（經 C）。

2. 問卷測量的信效度表現

有 4 位受訪者（社 C、政 A、政 B、心 A）表示其所屬領域，有些研究概念與變數的測量方式已有學界公認的測量方式，在判斷題項是否可用時，會觀察其與公認的測量方式是否一致：

例如說我們在講「政治信任」，我們這個領域已經發展出說我們會用四個題目來測量一個人的政治信任，就這四個題目。只要在這個領域，不用那四個題目，我們就認為是不對的（政 B）。

某些問卷改編自國外已經過信效度檢驗的問卷，但改編後與原本問卷不完全一致，這時候，學者就必須判斷這筆既有資料是否還具有良好的測量效果。

原本（註：國外發展的原始問卷）二十題到最後（註：國內的問卷）十題，只剩十題，那這樣我們用這個量表所得到的，還代表他一開始想要的問的構念嗎？我們在心理測量就要去問這一個基本的問題（心 B）。

除上述之外，心 A 與政 A 也會觀察資料的回應狀況，來評估問卷的測量品質，例如：

我覺得簡單的經驗法則是，如果有一個問題，你放在臺灣被測量，你看到它的變異很小，大家都集中在同意或不同意，滿意或不滿意，這就表示它的測量不好。它沒有變異（政 A）。

3. 問卷能提供的控制變數是否夠豐富

社 C、經 A 與經 C 均表示，若同一變數資料可同時在不同調查或數據庫獲得，就會選擇涵蓋較多變數的那份資料，因為資料集裡能提供的控制變數愈多，後續的推論統計分析就更有解釋力。例如：

有薪資調查的就有非常多，家庭收支、人力資源、受僱員工薪資調查都有。所以怎麼樣抉擇，就看我要研究的薪資之外，我其他想要控制的變數是什麼。因為受僱員工薪資調查沒有家人的資料，它就是從公司裡面去做調查，他只知道說你這個人在哪一家公司，但是你這個人有沒有結婚、教育程度、有沒有子女，在這個資料庫裡是看不到的。所以你如果想要知道說他有沒有結婚，對薪資有沒有影響，受僱員工薪資調查就不適合（經 C）。

4. 問卷題項的邏輯是否合理

社 A 在選擇資料時，會觀察問卷題目在題序安排、主題分布、或跳題順序的組織方面是否合理，一旦邏輯不佳，資料品質可能就會低落：

比如它問你是不是派遣員工……接下來會請派遣員工跳答，結果下一題又問說你是不是自營業者。因為說他的跳答邏輯沒有設計好，以致於有些人應該跳答又沒有跳答，所以整個後來的問卷資訊其實是有落差的（社 A）。

5. 多年期間卷題項的穩定程度

教 C 指出，如果取用的是多年期系列調查資料，歷年題項的穩定程度很重要，必須逐一核對每年題項是否一致，才能進行跨年比較分析。此外，雖然資料說明書可能有說明題項變動，但因說明的詳簡程度不一，為求慎重，研究者仍須耗費心力自行比對題項差異：

可能每一波調查題項的敘述方式有變，但都是在調查同一件事情，這時候我們就要做對照表……這其實是很費功夫的……每一個題項是什麼涵意，它（資料說明書）應該要寫清楚，如果是分年度的調查，第二年他們有修正了什麼，應該要有勘誤表或修正表讓我們對照，我們就不用再自己對了（教 C）。

(二) 評估樣本品質

有 11 位受訪者（社 A、社 B、社 C、政 A、教 A、教 B、教 C、經 A、經 B、經 C 與心 A）會觀察樣本規模及其代表性。樣本夠大才能反應個體差異，且樣本組成也必須吻合母體的分佈。

在作實證分析最怕的是樣本數太少，樣本數太少，代表性就不夠，而且在估計時偏誤性會比較高。所以我們在做實證分析，最少我們希望 100 筆以上，其實 100 筆 200 筆對我們來講都還不太夠……當我在做一些資料的篩選，比如說我們想要看男女差異，資料在做切割的時候，切割開資料數不會只剩下一點點？[又]比如說女性我們又再分成未婚、已婚，不斷切割的時候，[樣本必須]不至於到最後只剩下兩三個（經 C）。

[樣本]的抽樣分配，男生有多少女生有多少，教育程度有多少，大專教育百分比是多少人…有的還要跟臺灣人口作比對，才會具有代表性……像我問你，如果[樣本中]男生有三分之二，女生有三分之一，你覺得這個樣本可以用嗎？（經 B）

對部分教育學研究者而言，長期樣本的可得性，以及每筆樣本資料的確切可辨，會影響其能否取用資料。例如教 C 在觀察資料時，就會注意資料是否提供樣本 ID 號碼，因為缺了這個資訊，就無法判斷是否為追蹤資料：「要有 ID，你才能從 ID 去追蹤到底後續有沒有再去追蹤調查……所以我們要看資料裡面有沒有提供 ID 這一項。」

(三) 評估資料蒐集過程

即便問卷工具設計得當，樣本狀況也佳，研究者還會在意資料生產的過程，也就是資料蒐集的過程及其對資料品質的影響。由於資料蒐集是他人所執行，且在取用之時已無從查考，受訪者多半仰賴資料釋出機構的權威度來拿捏資料的品質，多數受訪者（如社 A、社 B、社 C、政 A、政 B、政 C、經 A、經 B、經 C、心 A 等）表示比較信任政府機關、學術機構或民間官方協會的資料，因為這些機構在資料蒐集過程中，比較有能力謹守抽樣與蒐集步驟的一致性。例如：

比如說[當]我們[抽樣]抽到這個人……那坦白講，你訪問不到他，你不能[隨便]換別人，從理論來說他就是被我們抽到的人，所以我們會要求訪員。一個有可信度的團隊，有經驗的，至少會

要求訪員去三次啦，例如這個就是一個評估的標準（政 B）。

相對而言，一旦有證據指涉到某資料的抽樣或蒐集過程不嚴謹，研究者可能就會放棄資料。例如經 B 表示曾見識過某些個別研究計畫由學生助理進行資料蒐集，問卷與抽樣品質皆不佳，但研究者也將錯就錯，從此他對個別研究釋出的資料，都較不具信心。政 C 也曾透過同儕口耳相傳，得知某機關的某特定調查研究計畫，其資料蒐集涉及舞弊，訪員會擅自修改無效樣本，從此，他就不再信任該計畫所釋出的任何資料：

訪員爲了 [達到] 成功的樣本數，他甚至後面的問題是自己勾，自己創造答案，那理論上這筆是無效樣本，那個資料不能納入資料庫裡。我是有聽說 [該調查] 一旦拿掉這些無效樣本，[樣本數就會不夠]……就算那個資料裡面有你可以用的部份，你也不太可能用它，你對他沒有信任感（政 C）。

另一個與資料蒐集品質相關的難題，是難以蒐集的資料，無論是難以填寫或在調查時缺乏數據者均屬之，這會形成資料集內的遺漏值，例如社 C 舉例：「像犯罪統計的，像國際犯罪統計年報，那個變動就很大，有很多遺漏值。」遺漏值一旦太多，就表示資料的可信度不高。但如果無法規避使用資料，他們就會嘗試用統計技術來校正資料品質，例如，國際性統計多會附上各個國家的統計品質評比，因此就將各國評比分數設為控制變數，藉以降低統計品質不一致的影響。

（四）評估資料的新穎度

資料新穎程度是否影響資料取用，見仁見智。理論上，資料新穎度的重要性應該是取決於問題意識，六位受訪者（教 C、政 A、政 B、政 C、經 C、心 A）也認同這個看法，然而，教 A 與經 A 卻指出，他們從自身的資料再用經驗中，感覺臺灣學界似乎比較要求資料新穎度，也比較會質疑過時資料的價值：

更新是很重要的東西，你做出來的報告可能會被質疑說時間過久。我還有幾篇文章是這樣被退稿的，臺灣學術界認爲教育問題是與時俱進的（教 A）。

[聽說有同儕] 投稿的時候都會被噓這件事（註：資料太舊），可

是好像在國外投稿就還好……因為經濟發展過了就過去了，當然是以新資料為主。或者會想要去更新資料，不會只看舊資料（經 A）。

（五）評估資料易得性

經 B 與經 C 指出，在有相通性質的資料可在不同地方取得時，會考慮的是資料方便取用性，例如：是否可一次下載多種資料或資料是否免費：

主計處其實是彙整所有部會的資料。所以與其你到各部會去找資料，不如直接到主計處他直接彙整好給你用（經 C）。

像有一項我們做在人力資源調查，剛開始是要給錢的啊……後來中研院開始做這件事，大概十年前開始做這件事，他就變免費的，變免費我就跟中研院拿就好（經 B）。

受訪者政 C 曾因為語言以及不熟悉國外學界，無法找到其感興趣的資料，而放棄原有的研究問題：

我對義大利的個案很感興趣，可是我對義大利學界不熟，我也不太清楚我要去哪裡取得義大利的調查資料，再加上不懂義大利文。種種限制之下我只好放棄這個主題（政 C）。

（六）評估分析結果的發表潛力

社 A、社 B、社 C 與教 A 等四位表示，能否從資料分析出具有發表價值的結果，為相當重要的因素。受訪者會希望研究分析的統計結果是顯著的、可解釋現實現象、或分析觀點具有新穎性。一旦對初步分析的結果不滿意，便會停止使用該資料：

沒有辦法產生有意義的研究我們就覺得不好用啊……如果從資料庫裡沒有辦法得到一些重要的新的知識的話，我覺得沒有太大用途啦。有些東西我們都已經知道啦，或是已經不需要再研究了（教 A）。

受訪者經 A 不太使用個別研究釋出的資料，理由就是認為個人蒐集的資料是在服務該學者特定的研究問題，那些資料釋出之時，便已

經不具有分析價值了：

他選的變數他自己喜歡的，他可能已經做了幾篇論文，那我們再去追這個資料是沒什麼意義的（經 A）。

社 B 會觀察前人基於該資料而發表的論文，從論文的品質也可判斷該資料的再用潛力：

其他人已發表的文章，也可以知道資料品質到哪裡……研究的可靠程度、能不能發表都會有關……從其他人發表的論文來看資料品質好不好，來判斷研究潛力（社 B）。

五、資料處理

受訪者獲得資料後，在進行正式分析之前，會先對資料進行描述統計與信效度檢驗，後續再做基本數據處理以及資料合併等處置，以下分別說明其目的與作法：

（一）進行描述統計與信效度檢驗

有九位受訪者（社 A、社 B、教 A、心 A、政 A、政 A、政 B、政 C 與經 C）表示，在下載資料後，會先對資料進行描述統計，以瞭解樣本與題項分布狀況，並可同時觀察資料是否有異常。此外，心 A、心 B、政 A 與政 C 等拿到外部資料時，通常都會自己再執行一次信效度檢驗，確認整體數據品質。但如果資料來源是已有信任感的數據庫，就不一定會再費事去檢驗：

真的在做研究時，當我們拿到一筆資料，而且我們信任說這個資料是來自於學術單位執行，而且大家都在用，應該有一定品質時，說真的我們也不太會再去信效度的檢驗（政 C）。

（二）基本數據校正處理

受訪者對數據進行的基本處理包含檢查數據錯誤、處理遺漏值，及變數轉換等，目的是將資料數據調整至堪用程度，並轉換為符合本次分析所需要的變數形式。以下簡述：

1. 檢查數據錯誤

社 A、社 C、政 A、政 C、教 B、經 C 與心 A 等會透過邏輯檢誤與次數分配來檢查數據是否有錯誤。邏輯檢誤是從題項與答覆之間的邏輯關係來檢查數據是否不合理；次數分配檢查則在觀察各題項的次數分配加總是否等同樣本數量，有無存在異常分布等。

比如說他說他沒工作，理論上接下來有關工作的訊息他都應該要跳答。可是結果他下一題問說，你在哪裡工作，他又回答，我在台中市工作，那這兩個是矛盾的（社 A）。
你寫語法進去看次數分配、最大值最小值，就知道資料對不對……檢查它的數值看跟原本樣本一不一樣（教 A）。

2. 處理遺漏值

社 B、社 C、教 A、教 C、經 B、經 C、政 C 等也提及遺漏值的問題，大致上可透過插補法、刪除樣本、使用替代值來處理。

- (1) 插補法：利用已知的數值計算函數，再用函數來推估遺漏的缺失值，這是一種常用的統計方法。教 A 說：「有時候你缺失值蠻多的時候，可以用插補，讓樣本不要大量流失。」
- (2) 刪除樣本：當遺漏值過多時，使用插補法並無太大意義，不如刪除樣本。
- (3) 使用替代值：經常使用國際資料的社 C 指出，如果某國家缺乏某特定年度的資料時，或可使用該國相近年度的數據來取代數據缺乏的那個年度，前提是兩個年度之間並無重大變化。

3. 變數轉換

多位受訪者（社 A、社 B、政 C、教 B、經 B、經 C 與政 C）提及，原始資料的變數測量方式不見得與他們的需求相符，為了進行他們想作的統計分析，可能會需要調整原有資料的編碼方式，或進行變數整併與對數轉換等。

- (1) 調整編碼方式：指將變數的編碼方式，調整為再用分析時想要的編碼方式，例如整併名義變數的類別：
我們在統計上不可能處理太多類別……它原本教育程度分成二十個類別，可是二十個類別對我們統計來說，它太細了……我們有時候會把它說變成三個類別或者五個類別，或者教育年數，幾乎是每個變數都要處理（社 A）。
- (2) 變數整併：為使資料變數能與受訪者欲測試的理論架構一致，可能需要將多個變數化約成一個單一變數：

我們在測量的時候，可能會單獨測你有沒有偏向哪一個政黨：國民黨、親民黨、民進黨……我就要整併兩個變數，比方說整併成強烈偏國民黨、中等偏國民黨、弱偏國民黨、強烈偏民進黨，用光譜的形式來處理（政 C）。

- (3) 對數轉換：社 C 指出，當觀察到資料分布有統計上的偏態時（意即資料分配狀況不是常態分配），便必須進行對數轉換，例如將數值轉換成自然對數，在後續統計分析上才較具有說服力。

（三）合併資料集或填補數據

當受訪者所用的資料是多年追蹤調查或跨族群調查時，可能需要合併多個資料集。此外，當資料集缺乏受訪者需要的變數，或可由其他外部資料來尋找替代變數，或自己執行一個補充調查來補入變數。

1. 合併多年期或跨族群的資料

政 A、社 A、社 B、教 A、教 B 與教 C 等都曾進行兩組以上的資料合併，在資料合併時，最重要的就是比對樣本編號，確認同一樣本或有關聯的樣本能夠準確對應，資料才具有分析效力：

如果是一個多次調查的資料，第一你要先串聯不同的資料檔，你不可能一次拿到好幾年的資料，你要先比對資料樣本的 ID，再去合併資料。比如說第一年、第三年、第五年，你每一年的資料都要幫作者做一下確認（社 A）。

編碼裡面有一個 ID，家長有一個 ID，學生有一個 ID，兩個 ID 要一樣才能 match……如果你要做不同階層的分析，你就要把那個碼整合起來。所以這個時候就開始要整合（教 A）。

2. 從外部來源尋求控制變數

經濟領域的三個受訪者表示，在進行分析時，除了使用既有的資料集之外，往往還需要從外部來源取得社經環境變化的資料，作為控制變數，如此分析才會合理：

比如說你 [研究離職率，你還需要的外部] 資料是 2007、2008、2009、2010 那時候的總體的經濟狀況，像 GDP 或是失業率，你可以把它拿來控制當時的經濟環境，會不會影響你個人的離職率。因為離職率並不是只有個人因素，它還有外在因素（經 A）。

3. 補充原資料集不足的變數

當既有資料缺乏研究分析所需要的變數時，可以透過組合變數、使用工具變數，或自行執行一個補充研究來填補變數。

- (1) 組合變數：當教 B 發現既有資料的問卷題項與其分析架構不完全吻合時，會先檢視既有的問卷題項，尋找意旨相近、組合後足以替代的題項。如有勉能替代且能通過信效度檢驗者，即可用之：

我做的東西是學校歸屬感，國外是用一個 21 題的問卷……我剛說的那批研究者，發現學校歸屬感對學生成就是有影響的，所以他們就用美國各種資料庫裡面，有用到學校歸屬感這種東西的變數去拼，拼出一個長得像學校歸屬感的量表出來，到目前為止我用的方式也是這樣（教 B）。

- (2) 使用工具變數：當資料內沒有自己想用的變數，可嘗試使用其他相關變數來取代，是為工具變數。既有資料中的變數能否當成工具變數來使用，則要看文獻或常理邏輯能否充分支持兩變數間的合理替代關係：

比如說，我們想要知道這個家庭它休閒時間多長，但是家庭收支資料庫就沒有工時，就沒有辦法推出休閒時間，所以我們可能就用休閒支出做休閒時間的推測（經 C）。

- (3) 執行補充研究來填補變數：如果前述方法均不可行，研究者要不就放棄該筆數據，要不就是執行一個補充性的研究來填補資料：

比如說，那篇文章主要是看你的家庭背景會不會影響你在進入市場就業時候的風險選擇……但是資料上就會比較缺乏是這個人長時間的調查我們那時候的作法是，從幼稚園到研究生，從 5 歲一直到 25 歲的人，每一個年齡層都做了 100 個樣本……所以當我們湊滿 5 歲到 25 歲各 100 個樣本每個人的家庭背景狀況、風險態度之後，再對應到我們取用的那個資料庫去（經 C）。

伍、討論

前文已依序呈現本文受訪者的資料再用動機、獲知管道、資料評估策略、以及分析前的資料處理行為。研究結果或與前人文獻呼應，或有前人未及之新論點。

首先，在資料再用動機或理由方面，Yoon（2015）指出對部分社

會科學領域而言，資料收集非個人之力所能，這點與本文發現吻合。但相對而言，本文受訪者較少認為資料再用是因為既有資料已夠充足，無須再作收集，反之，本文受訪者在再用資料時，心態通常是「不得不為」，因為取用既有資料是唯一的途徑（例如：經濟學或是政治學中的選舉研究），或「相對可為」（例如運用大規模教育統計，資料規模大且有公信力，比較能說服審稿者，也可免除倫理審查麻煩等）。對某些學科或主題來說，例如總體經濟、政治學中的民意與選舉研究、人口學等，數據資料幾乎不可能自行收集，仰賴政府或大型科研調查機構釋出的數據，幾乎是領域慣習，這也說明了為何多數受訪者曾取用的資料類型，幾乎都是系列調查與業務統計。

社會科學研究者運用前人個別研究產製的資料，來探索新議題或處理前人未及的議題，雖然仍有可能，但都是少數，而且這些再用者與資料原始產製者，可能都必須要具有一定的社會關係或先備的合作關係。誠如受訪者經 A 指出，社會科學研究者其實很少能取用前人個別研究的數據，再進行有意義的二次分析，因為能做的分析，原作者通常自己都做完了，後人很難透過再用分析，提出具原創性的發現。

在獲知管道上，本文的發現與前人研究也大致相符：人際管道、數據庫、期刊論文、與資料說明書等，均能揭示資料存在（Faniel & Jacobsen, 2010b; Sands, Borgman, Wynholds, & Traweek, 2012; Kriesberg, Frank, Faniel, & Yakel, 2013; Daniels, 2014）。但對本文受訪者而言，以人次論，最多是透過學術文獻來獲知，其次為人際網絡以及政府與學術機構之網站。此外，也有幾位受訪者指出，學會或資料機構所舉辦的推廣活動或電子報，對他們獲知資料存在是有幫助的，顯示這類活動值得繼續發展與鼓勵。

在資料評估策略上，本文分析受訪者的評估行動，指出受訪者可能會透過觀察問卷工具品質、樣本品質、資料蒐集過程、資料新穎度、易得性、與分析發表潛力等，來決定是否取用資料。與前述文獻分析所歸納的四大影響因子交叉比對，可得關聯如表 3。本文受訪者所提多種行動策略皆聚焦於正確性／可信性，可用性項目則次之，也隱然呼應前人發現的重要性順序。

表 3.
資料評估行為與資料選擇影響因子之對照

資料評估行動策略	正確性 可信性	可用性	可及性	相關性
● 評估問卷工具品質				
○ 問卷能否回應再用者關心的問題				✓
○ 問卷測量的信效度表現		✓		
○ 問卷提供的控制變數是否夠豐富		✓		
○ 問卷題項的邏輯是否合理	✓			
○ 多年期間問卷題項的穩定程度	✓	✓		
● 評估樣本品質	✓	✓		
● 評估資料蒐集過程	✓			
● 評估資料新穎度				✓
● 評估資料易得性			✓	
● 評估分析結果的發表潛力				✓

在資料處理行為部分，文獻中則未見前人研究。本文透過開放提問來了解社會科學學者的資料處理行為。值得注意的是，社會科學跟自然科學一樣，都常使用兩筆以上的資料集，但是社會與自然學者操作多筆資料集的方式是否相同或相似，則未可知。在社會科學研究中，學者經常關心概念性的變數，但概念須由測量指標與問卷題項來代表，這與自然科學甚不相同。因此，不同資料集的合併使用是否合理，就是一大問題，舉例來說，長期社會調查的題項如果有異動，那麼不同時期的資料是否能夠合理整併，是否都代表同一件事，這就考驗學者的判斷。另外，由於社會調查並不是發生在經過控制的環境，因此在分析時，學者也常需要控制外部變數來輔助觀察，但分析時需要哪些控制變數、如何取得，這些都突顯出社會與自然科學研究的差異。社會科學與自然科學對量化資料的處理與運用行為，應是再用行為研究的核心，值得未來再深入探索。

陸、結論

本文透過訪談 14 位社會科學學者來瞭解其資料再用動機或理由、獲知管道、資料評估策略、以及正式分析前的資料處理行為。歸納而言，在再用動機方面，學者取用既有資料，理由可能是因為無法自行蒐集大規模或長期性資料、取用來自權威單位的資料比較具有公信力、取用既有資料可免除研究倫理審查的煩擾、可探索潛在研究題目或能拓展既存的研究議題，同時，學科領域文化對資料再用也可能產生鼓勵或限制的作用。在獲知管道方面，本研究則辨識出五種常用管道，包含學術文獻、同儕與指導教授、政府與學術機構網站、學會或調查機構的推廣活動、紙本統計資料等。

在資料評估方面，本研究辨析出六種社會科學學者常用的評估策略，分別是評估問卷工具品質、評估樣本品質、評估資料蒐集過程、評估資料新穎度、評估資料易得性、以及評估分析結果的發表潛力。其中，在評估問卷工具品質上，受訪者共提出了五種不同觀察角度，分別是問卷能否回答再用所關心的問題、原始問卷在測量上的信效度表現、問卷能提供的控制變數豐富程度、問卷題項合理程度、多年期間問卷題項的穩定程度等。這些評估策略的訴求，均能呼應某種資料評估的影響因子，例如資料的正確性／可信性、可用性、可及性、與相關性等。

最後，在資料處理行為部分，本文辨析學者在分析既有資料前，必須進行的資料耙梳與整理工作，發現他們首先必須進行描述統計與信效度檢驗，確認資料可用，接著才開始進行基本數據校正處理，例如檢查數據錯誤、處理遺漏值、進行變數轉換等。後續如有必要，還須比對並合併多組資料，當資料集欠缺研究者所需變數時，也會由其他外部資源尋求控制變數，或透過組合變數與工具變數等來補充原資料集之不足。

上述發現的貢獻，就實務而言，能使我們更深入瞭解社會科學研究者的量化數據再用行為與需求，對資料產製者、分享者或資料典藏與傳播單位，都應有所助益；對研究而言，本文以質化訪談的方式探知社會科學研究者的資料品質評估策略，並具體描繪了分析前的資料處理行為，這些都是前人文獻未曾探討過的資料使用行為。未來研究或可奠基於本文發現，深入且大規模地探索自然與社會科學領域之間，或是社會科學各子領域內，在資料再用動機、獲知管道、評估策略與資料處理等方面行為，是否有系統性差異，或具領域獨特性，如此應

更能深化資料行為研究，對資料分享與資料服務實務工作應有所助益。

致謝

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。

參考書目

- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059-1078.
- Butler, D. (2006). Mashups mix data into global service: Is this the future for scientific analysis? *Nature*, 439(7072), 6-7.
- Carlson, S., & Anderson, B. (2007). What are data? The many kinds of data and their implications for data re-use. *Journal of Computer-Mediated Communication*, 12, 635-651.
- Curry, R. G. (2015). *Beyond “data thrifting”: An investigation of factors influencing research data reuse in the social sciences* (Doctoral dissertation). Retrieved from <http://surface.syr.edu/etd/266/>
- Curry, R. G., & Qin, J. (2014). Towards a model for research data reuse behavior. *Proceedings of the American Society for Information Science and Technology*, 51(1), 1-4. doi: 10.1002/meet.2014.14505101072
- Daniels, M. G. (2014). *Data reuse in museum contexts: Experience of archaeologists and botanist* (Doctoral dissertation). Information and Library Studies in The University of Michigan, Michigan.
- Faniel, I. M., & Jacobsen, T. E. (2010a). *Reusing scientific data: A research framework*. Paper presented at CSCW 2010 Workshop on “the changing dynamics of scientific collaboration”, Savannah, GA.
- Faniel, I. M., & Jacobsen, T. E. (2010b). Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues’ data. *Compute Supported Cooperative Work*, 19, 355-375.
- Faniel, I. M., Barrera-Gomez, J., Kriesberg, A., & Yakel, E. (2013). A comparative study of data reuse among quantitative social scientists and archaeologists. *iConference 2013 Proceedings*, 797-800. doi:

10.9776/13391

- Faniel, I. M., Kansa, E., Kansa, S. W., Barrera-Gomez, J., & Yakel, E. (2013). The challenges of digging data: A study of context in archaeological data reuse. In *JCDL 2013 Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp.295-304). New York, NY: ACM.
- Faniel, I. M., Kriesberg, A., & Yakel, E. (in press). Social scientists' satisfaction with data reuse. *Journal of the Association for Information Science and Technology*. doi:10.1002/asi.23480
- Faniel, I. M., Kriesberg, A., Yakel, E. (2012). Data reuse and sense making among novice social scientists. *Proceedings of the American Society for Information Science and Technology*, 49(1), 1-10. doi: 10.1002/meet.14504901068
- Fear, K. M. (2013). *Measuring and anticipating the impact of data reuse* (Doctoral dissertation). Information and Library Studies in The University of Michigan, Michigan.
- Fienberg, S. E., Martin, M. E., & Straf, M. L. (Eds.). (1985). *Sharing research data*. Washington, DC: National Academies Press.
- Fischer, B. A., & Zigmond, M. J. (2010). The essential nature of sharing in science. *Science and Engineering Ethics*, 16(4), 783-799.
- Fjordback Søndergaard, T., Andersen, J., & Hjørland, B. (2003). Documents and the communication of scientific and scholarly information: Revising and updating the UNISIST model. *Journal of documentation*, 59(3), 278-320.
- Gamble, M., & Goble, C. (2011). *Quality, trust, and utility of scientific data on the web: Towards a joint model*. Paper presented at WebSci'11 Proceedings of the 3rd International Web Science Conference. New York, NY. doi: 10.1145/2527031.2527048
- Huang, H., Stvilia, B., Jorgensen, C., & Bass, H. W. (2012). Prioritization of data quality dimensions and skills requirements in genome annotation work. *Journal of American Society of Information Science and Technology*, 63(1), 195-207.
- 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.
- Kriesberg, A., Frank, R. D., Faniel, I. M., & Yakel, E. (2013). The role of

- data reuse in the apprenticeship process. *Proceedings of the American Society for Information Science and Technology*, 50(1), 1-10. doi: 10.1002/meet.14505001051
- Kruse, F., & Thestrup, J. B. (2014). Research libraries' new role in research data management, current trends and visions in Denmark. *Liber Quarterly: The Journal of European Research Libraries*, 23(4), 310-335.
- Kvalheim, V., Kvamme, T., & Norwegian Social Science Data Services. (2014). *Policies for sharing research data in social sciences and humanities: A survey about research funders' data policies*. Retrieved from http://ifdo.org/wordpress/wp-content/uploads/2014/04/ifdo_fact.pdf
- Lagoze, C., & Velden, T. (2009). Communicating chemistry. *Nature Chemistry*, 1, 673-678.
- Mauthner, N. S., & Parry, O. (2009). Qualitative data preservation and sharing in the social sciences: On whose philosophical terms? *Australian Journal of Social Issues*, 44(3), 291-307.
- Mayernik, M. S., Wallis, J. C., Pepe, A., & Borgman, C. L. (2008). *Whose data do you trust? Integrity issues in the preservation of scientific data*. Paper presented at iConference 2008. Retrieved from <https://www.ideals.illinois.edu/handle/2142/15119>
- Murillo, A. P. (2014). Examining data sharing and data reuse in the dataone environment. *Proceedings of the American Society for Information Science and Technology*, 51(1), 1-5. doi: 10.1002/meet.2014.14505101155
- 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.
- Niu, J. (2009). *Perceived documentation quality of social science data* (Doctoral dissertation). Information and Library Studies in The University of Michigan, Michigan.
- Niu, J., & Hedstrom, M. (2009). *Documentation evaluation model for social science data: An empirical test*. Paper presented at DigCCurr 2009

- Conference. Chapel Hill: NC
- Piwowar, H. A. (2008). Proposed foundations for evaluation data sharing and reuse in the biomedical literature. *Bulletin of IEEE Technical Committee on Digital Library*, 4(2). Retrieved from <http://www.ieeetcdl.org/Bulletin/v4n2/piwowar/piwowar.html>
- Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation. *PeerJ*, e175. doi: 10.7717/peerj.175
- Read, E. J. (2007). Data services in academic libraries: Assessing needs and promoting services. *Reference & User Services Quarterly*, 46(3), 61-75.
- Sands, A., Borgman, C. L., Wynholds, L., & Traweek, S. (2012). Follow the data: How astronomers use and reuse data. *Proceedings of the American Society for Information Science and Technology*, 49(1). doi: 10.1002/meet.14504901341
- Santer, B. D., Wigley, T. M. L., & Taylor, K. E. (2011). The reproducibility of observational estimates of surface and atmospheric temperature change. *Science*, 334(6060), 1232-1233.
- Strauss, A. & Corbin, J. (2001)。質性研究入門：紮根理論研究方法 (*Basics of qualitative research : Techniques and procedures for developing grounded theory, 2nd ed*，吳芝儀、廖梅花譯)。嘉義市：濤石文化。(原著出版於1998年)
- Stvilia, B., Hinnant, C. C., Wu, Shuheng, Worrall, A., Lee, D. J., Burnett, K., Burnett, G., Kazmer, M., & Marty, P. F. (2015). Research project tasks, data, and perceptions of data quality in a condensed matter physics community. *Journal of the Association for Information Science and Technology*, 66(2), 246-263.
- 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. Retrieved from <http://www.qualitative-research.net/index.php/fqs/article/view/499/1074#g3>
- Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS ONE*, 8(7), e67332. doi: 10.1371/journal.pone.0067332

- Whitlock, M.C. (2011). Data archiving in ecology and evolution: Best practices. *Trends in Ecology & Evolution*, 26(2), 61-65.
- Wynholds, L., Fearson, D. S. Jr., Borgman, C. L., & Traweek, S. (2011). When use cases are not useful: Data practices, astronomy, and digital libraries. *JCDL' 11 Proceedings of the 11th annual international ACM/IEEE joint conference on digital libraries*, 383-386. doi: 10.1145/1998076.1998146
- Xia, J., & Liu, Y. (2013). Usage patterns of open genomic data. *College & Research Library*, 74(2), 195-207
- Xia, J., & Wang, M. (2014). Competencies and responsibilities of social science data librarians: An analysis of job descriptions. *College & Research Libraries*, 75(3), 362-388.
- Yoon, A. (2014). “Making a square fit into a circle”: Researchers’ experiences reusing qualitative data. *Proceedings of the American Society for Information Science and Technology*, 51(1), 1-4.
- Yoon, A. (2015). *Data reuse and users’ trust judgments: Toward trusted data curation* (Doctoral dissertation) . Retrieved from <https://cdr.lib.unc.edu/record/uuid:2c2268b3-88cf-4397-b038-b39e88f80d83>
- Zenk-Moltgen, W., & Lepthien, G. (2014). Data sharing in sociology journals. *Online Information Review*, 38(6), 709-722.
- Zhang, J. (2011). *Data use and access behavior eScience – Exploring data practices in the new data-intensive science paradigm* (Doctoral dissertation). Retrieved from <https://idea.library.drexel.edu/islandora/object/idea%3A3543>
- 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, 1-14. doi: 10.1007/s11192-010-0318-1
- Zimmerman, A. S. (2008). New knowledge from old data: The role of standards in the sharing and reuse of ecological data. *Science, Technology, & Human Values*, 33(5), 631-652.

Data Reuse Behavior among Taiwan Social Scientists

Chi-Shiou Lin

Associate Professor
Department of Library and Information Science,
National Taiwan University

Ching-Yi Lai

Graduate Student
Graduate Institute of Library and Information Science,
National Taiwan University

Introduction

This study investigated how social scientists in Taiwan use existing quantitative data for original or secondary analyses. In recent years, data sharing has been a salient topic in academia. Empirical studies on how scientists use existing data have also emerged. However, data reuse in the social sciences is less studied than in natural and applied sciences. This study focuses on why and how social scientists use existing quantitative data in their research, including survey data collected by government agencies or research institutions, all kinds of business data from the public and private sectors, and data obtained from other researchers. Specifically, this study sought to answer the following research questions:

1. What are the motivations for reuse of existing quantitative data by social scientists?
2. How do social scientists become aware of existing data?
3. How do social scientists evaluate existing data for possible reuses?
4. What data processing tasks are required in the reuse of existing data?

Research Method

This study employed semi-structured in-depth interviews to

understand how social scientists reuse existing data. Purposive sampling was used to recruit participants from five larger social sciences divisions as defined by the Ministry of Science and Technology, including sociology, political science, education, economics, and psychology. The researchers first examined the journals indexed in the 2015 TSSCI to identify recent published empirical research papers based on quantitative data reuse. The researchers then invited the first authors of the papers to participate in a one-on-one interview. For each of the first four divisions, three participants were successfully interviewed. Only two psychology researchers were interviewed due to the very small amount of published research from the source used. A total of fourteen participants reported on their motivations for and experiences with quantitative reuses. Each interview took 1-2 hours. All interviews were recorded and transcribed verbatim for further analyses. As can be seen in Table 1, all of the participants had experience with survey data reuses; only four of them had employed business data in their previous research; and very few had used data generated from previous research.

Table 1.
Participant Profiles

ID	Division	Seniority	Reuse by Data Type		
			Survey data	Business data	Previous research data
SocA	Sociology	Over 10 years	✓		
SocB	Sociology	7 years	✓		
SocC	Sociology	Over 20 years	✓	✓	
PoliA	Political Science	Over 10 years	✓		
PoliB	Political Science	15 years	✓		
PoliC	Political Science	Over 10 years	✓		
EduA	Education	Over 10 years	✓		
EduB	Education	5 years	✓		
EduC	Education	4 years	✓		
EcoA	Economics	20 years	✓	✓	
EcoB	Economics	Over 10 years	✓	✓	✓

EcoC	Economics	18 years	✓	✓	
PsyA	Psychology	3 years	✓		✓
PsyB	Psychology	Over 10 years	✓		

Note. Survey data are often continuous data released from government agencies or public/private research institutions. Business data refer to the compilation of statistics of regular business activities of the public and private sectors. Previous research data are those from previous studies conducted by other researchers.

Study Findings

Analyses showed that, in regards to motivations, participants re-used existing data for the following reasons: (1) they were unable to collect large-scale or long-term data on their own; (2) existing data were perceived as having higher credibility especially those released by government or authoritative sources; (3) using existing data may free them from the burden of IRB reviews; (4) they may explore potential research questions with existing data without having to worry about the cost of data collection and possible failed investment; (5) they may continue to explore a particular stream of research inquiries with existing data; and (6) organizational culture may encourage or discourage the reuse of data.

Participants relied on five different channels to find existing data: (1) research literature; (2) peers and advisors; (3) government agencies or academic institutions' websites; (4) promotional activities of scholarly associations and survey institutes; and (5) statistical publications.

Participants reported six evaluation strategies for deciding whether to reuse the data: (1) evaluating survey instrument quality; (2) evaluating sample quality; (3) evaluating data collection procedures; (4) evaluating timeliness of the data; (5) evaluating ease and cost of access to the data; and (6) evaluating possible acceptance of study results based on the data. With the first evaluation item, i.e., evaluation of survey instrument quality, participants further reported five points of observation: (1) degree of topic match between the inquiry and the questionnaire content; (2) observed validity and reliability performance of the survey questionnaire; (3) whether the questionnaire offers rich control variables; (4) whether the questions are

organized in logical order; and (5) the stability of questionnaire items over time in chronological surveys.

Finally, in data processing, participants first inspected the descriptive statistics and validity and reliability of the dataset and then proceeded to correct, amend, or convert the data. They might also combine multiple datasets and fill in further needed variables from external sources or fill them in by using other techniques.

Discussion and Conclusion

The findings of this study were consistent with Yoon (2015) in that data reuses in social sciences were largely due to the inability of an individual scientist to collect large-scale data. The findings on the information channels used to become aware or stay alert of existing data were also consistent with previous research findings (Daniels, 2014; Faniel & Jacobsen, 2010b; Kriesberg, Frank, Faniel & Yakel, 2013; Sands, Borgman, Wynholds, & Traweek, 2012). For certain subject disciplines (e.g., economics) and some particular research domains (e.g., election research in political science), the use of existing data released from governments and reputed institutions is inevitable and taken-for-granted. Very few social scientists had used research data collected by other researchers for a particular research project unless social ties existed between the data sharer and the data reuser.

The researchers further mapped data evaluation strategies as identified in this study to the factors influencing data reuse reported in previous research. It helps to articulate the relationships between the behaviors and their goals.

Table 3.
Mapping data evaluation strategies with factors including data reuses

Data evaluation strategies	Accuracy; credibility	Usability	Accessibility	Relevance
● Evaluating survey instrument quality				
○ Topic match				✓
○ Perceived validity and reliability		✓		
○ Rich control variables		✓		
○ Logical sequence of questionnaire items	✓			
○ Stability of questionnaire items in chronological surveys	✓	✓		
● Evaluating sample quality	✓	✓		
● Evaluating data collection procedures	✓			
● Evaluating timeliness of data				✓
● Evaluating ease and cost of access to data			✓	
● Evaluating possible acceptance of study findings based on the data				✓

Another contribution of this study is about the articulation of data processing behavior prior to actual data reuse. In social sciences research, scholars are often concerned with “concepts”. In data processing and cleaning, data reusers often have to face the problems and challenges in combining and merging data from different sources to form a coherent and analyzable set of data. They also need a lot of control variables to arrive at more reasonable and credible analyses across the datasets. Those challenges distinguish data reuse behaviors and requirements of social scientists from

those of natural and applied scientists and await future investigation.

ROMANZIED & TRANSLATED REFERENCE FOR ORIGINALTEXT

- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059-1078.
- Butler, D. (2006). Mashups mix data into global service: Is this the future for scientific analysis? *Nature*, 439(7072), 6-7.
- Carlson, S., & Anderson, B. (2007). What are data? The many kinds of data and their implications for data re-use. *Journal of Computer-Mediated Communication*, 12, 635-651.
- Curry, R. G. (2015). *Beyond "data thrifting": An investigation of factors influencing research data reuse in the social sciences* (Doctoral dissertation). Retrieved from <http://surface.syr.edu/etd/266/>
- Curry, R. G., & Qin, J. (2014). Towards a model for research data reuse behavior. *Proceedings of the American Society for Information Science and Technology*, 51(1), 1-4. doi: 10.1002/meet.2014.14505101072
- Daniels, M. G. (2014). *Data reuse in museum contexts: Experience of archaeologists and botanist* (Doctoral dissertation). Information and Library Studies in The University of Michigan, Michigan.
- Faniel, I. M., & Jacobsen, T. E. (2010a). *Reusing scientific data: A research framework*. Paper presented at CSCW 2010 Workshop on "the changing dynamics of scientific collaboration", Savannah, GA.
- Faniel, I. M., & Jacobsen, T. E. (2010b). Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data. *Compute Supported Cooperative Work*, 19, 355-375.
- Faniel, I. M., Barrera-Gomez, J., Kriesberg, A., & Yakel, E. (2013). A comparative study of data reuse among quantitative social scientists and archaeologists. *iConference 2013 Proceedings*, 797-800. doi: 10.9776/13391
- Faniel, I. M., Kansa, E., Kansa, S. W., Barrera-Gomez, J., & Yakel, E. (2013). The challenges of digging data: A study of context in archaeological data reuse. In *JCDL 2013 Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp.295-304). New York, NY: ACM.
- Faniel, I. M., Kriesberg, A., & Yakel, E. (in press). Social scientists'

- satisfaction with data reuse. *Journal of the Association for Information Science and Technology*. doi:10.1002/asi.23480
- Faniel, I. M., Kriesberg, A., Yakel, E. (2012). Data reuse and sense making among novice social scientists. *Proceedings of the American Society for Information Science and Technology*, 49(1), 1-10. doi: 10.1002/meet.14504901068
- Fear, K. M. (2013). *Measuring and anticipating the impact of data reuse* (Doctoral dissertation). Information and Library Studies in The University of Michigan, Michigan.
- Fienberg, S. E., Martin, M. E., & Straf, M. L. (Eds.). (1985). *Sharing research data*. Washington, DC: National Academies Press.
- Fischer, B. A., & Zigmond, M. J. (2010). The essential nature of sharing in science. *Science and Engineering Ethics*, 16(4), 783-799.
- Fjordback Søndergaard, T., Andersen, J., & Hjørland, B. (2003). Documents and the communication of scientific and scholarly information: Revising and updating the UNISIST model. *Journal of documentation*, 59(3), 278-320.
- Gamble, M., & Goble, C. (2011). *Quality, trust, and utility of scientific data on the web: Towards a joint model*. Paper presented at WebSci'11 Proceedings of the 3rd International Web Science Conference. New York, NY. doi: 10.1145/2527031.2527048
- Huang, H., Stvilia, B., Jorgensen, C., & Bass, H. W. (2012). Prioritization of data quality dimensions and skills requirements in genome annotation work. *Journal of American Society of Information Science and Technology*, 63(1), 195-207.
- 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.
- Kriesberg, A., Frank, R. D., Faniel, I. M., & Yakel, E. (2013). The role of data reuse in the apprenticeship process. *Proceedings of the American Society for Information Science and Technology*, 50(1), 1-10. doi: 10.1002/meet.14505001051
- Kruse, F., & Thestrup, J. B. (2014). Research libraries' new role in research data management, current trends and visions in Denmark. *Liber Quarterly: The Journal of European Research Libraries*, 23(4), 310-

335.

- Kvalheim, V., Kvamme, T., & Norwegian Social Science Data Services. (2014). *Policies for sharing research data in social sciences and humanities: A survey about research funders' data policies*. Retrieved from http://ifdo.org/wordpress/wp-content/uploads/2014/04/ifdo_fact.pdf
- Lagoze, C., & Velden, T. (2009). Communicating chemistry. *Nature Chemistry, 1*, 673-678.
- Mauthner, N. S., & Parry, O. (2009). Qualitative data preservation and sharing in the social sciences: On whose philosophical terms? *Australian Journal of Social Issues, 44*(3), 291-307.
- Mayernik, M. S., Wallis, J. C., Pepe, A., & Borgman, C. L. (2008). *Whose data do you trust? Integrity issues in the preservation of scientific data*. Paper presented at iConference 2008. Retrieved from <https://www.ideals.illinois.edu/handle/2142/15119>
- Murillo, A. P. (2014). Examining data sharing and data reuse in the dataone environment. *Proceedings of the American Society for Information Science and Technology, 51*(1), 1-5. doi: 10.1002/meet.2014.14505101155
- 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.
- Niu, J. (2009). *Perceived documentation quality of social science data* (Doctoral dissertation). Information and Library Studies in The University of Michigan, Michigan.
- Niu, J., & Hedstrom, M. (2009). *Documentation evaluation model for social science data: An empirical test*. Paper presented at DigCCurr 2009 Conference. Chapel Hill: NC
- Piwowar, H. A. (2008). Proposed foundations for evaluation data sharing and reuse in the biomedical literature. *Bulletin of IEEE Technical Committee on Digital Library, 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. *PeerJ*, e175. doi: 10.7717/peerj.175
- Read, E. J. (2007). Data services in academic libraries: Assessing needs and promoting services. *Reference & User Services Quarterly*, 46(3), 61-75.
- Sands, A., Borgman, C. L., Wynholds, L., & Traweek, S. (2012). Follow the data: How astronomers use and reuse data. *Proceedings of the American Society for Information Science and Technology*, 49(1). doi: 10.1002/meet.14504901341
- Santer, B. D., Wigley, T. M. L., & Taylor, K. E. (2011). The reproducibility of observational estimates of surface and atmospheric temperature change. *Science*, 334(6060), 1232-1233.
- Science*, 334(6060), 1232-1233.
- Strauss, A. & Corbin, J. (2001) 。質性研究入門：紮根理論研究方法 (*Basics of qualitative research : Techniques and procedures for developing grounded theory, 2nd ed* , 吳芝儀、廖梅花譯) 。嘉義市：濤石文化。(原著出版於1998年) 【Strauss, A. & Corbin, J. (2001). *Basics of qualitative research : Techniques and procedures for developing grounded theory* (2nd ed. Zhi-Yi Wu & Mei-Hua Liao, trans.). Chiayi: Tao Shi Wen Hua】
- Stvilia, B., Hinnant, C. C., Wu, Shuheng, Worrall, A., Lee, D. J., Burnett, K., Burnett, G., Kazmer, M., & Marty, P. F. (2015). Research project tasks, data, and perceptions of data quality in a condensed matter physics community. *Journal of the Association for Information Science and Technology*, 66(2), 246-263.
- 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. Retrieved from <http://www.qualitative-research.net/index.php/fqs/article/view/499/1074#g3>
- Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share will anyone use them? Data sharing and reuse in the long tail of science and technology. *PloS ONE*, 8(7), e67332. doi: 10.1371/journal.pone.0067332
- Whitlock, M.C. (2011). Data archiving in ecology and evolution: Best practices. *Trends in Ecology & Evolution*, 26(2), 61-65.

- Wynholds, L., Fearson, D. S. Jr., Borgman, C. L., & Traweek, S. (2011). When use cases are not useful: Data practices, astronomy, and digital libraries. *JCDL' 11 Proceedings of the 11th annual international ACM/IEEE joint conference on digital libraries*, 383-386. doi: 10.1145/1998076.1998146
- Xia, J., & Liu, Y. (2013). Usage patterns of open genomic data. *College & Research Library*, 74(2), 195-207
- Xia, J., & Wang, M. (2014). Competencies and responsibilities of social science data librarians: An analysis of job descriptions. *College & Research Libraries*, 75(3), 362-388.
- Yoon, A. (2014). "Making a square fit into a circle": Researchers' experiences reusing qualitative data. *Proceedings of the American Society for Information Science and Technology*, 51(1), 1-4.
- Yoon, A. (2015). *Data reuse and users' trust judgments: Toward trusted data curation* (Doctoral dissertation) . Retrieved from <https://cdr.lib.unc.edu/record/uuid:2c2268b3-88cf-4397-b038-b39e88f80d83>
- Zenk-Moltgen, W., & Lepthien, G. (2014). Data sharing in sociology journals. *Online Information Review*, 38(6), 709-722.
- Zhang, J. (2011). *Data use and access behavior eScience – Exploring data practices in the new data-intensive science paradigm* (Doctoral dissertation). Retrieved from <https://idea.library.drexel.edu/islandora/object/idea%3A3543>
- 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, 1-14. doi: 10.1007/s11192-010-0318-1
- Zimmerman, A. S. (2008). New knowledge from old data: The role of standards in the sharing and reuse of ecological data. *Science, Technology, & Human Values*, 33(5), 631-652.