

數位工具與創新思維 ——社會科學國際資料庫應用工作坊活動紀要

撰文／徐靖淳（國立政治大學政治學系碩士生）、蔡維廷（國立政治大學政治學系碩士生）



國際資料庫是社會科學研究中相當關鍵的研究素材。透過跨國資料的分析，得以檢驗理論的適用範圍，並比較台灣與其他國家的異同，拓展我們對於社會科學理論的認識。鑒於國際資料庫的重要性，人社中心在2022年8月26日於國立臺灣大學集思臺大會議中心拉斐爾廳舉辦「社會科學國際資料庫應用工作坊」，依據資料的分析單位與運用方式規劃五大主題，邀請10位來自政治學、社會學、法律學、教育學、會計學、心理學等不同背景的學者，共同分享各自領域中被廣泛運用的國際資料庫。希望透過多元領域的資訊分享，促進跨領域的合作，併發全新的研究構想。

首先，人社中心的主任何明修教授進行開幕致詞，介紹人社中心的目標與業務內容，分別就個別研究者、學術社群、系所、出版社等不同的研究單位，說明人社中心所能提供的協助，並鼓勵與會者積極申請補助、參與人社中心舉辦的學術活動。接著，由中央研究院政治學研究所的吳文欽副研究員，分享籌備工作坊的過程、規劃各場次的目的，以及概述工作坊進行的流程。

第一場次的主題是「國家層次資料 (I)」，由國立政治大學國家發展研究所的蘇昱璇助理教授、國立成功大學政治學系的王奕婷副教授，以及國立政治大學政治學系的蘇彥斌教授擔任講者。蘇昱璇助理教授介紹的資料庫是「World Bank Indicators」以及「World Bank Open Data」，並表示如何運用這兩個資料庫有助於跨國比較，尤其是瞭解發展中國家的狀況。接著，蘇昱璇助理教授概述資料庫的基本資訊，並以多篇期刊論文為例，說明使用前述資料庫的方式，包括作為國家的背景介紹、或者用來比較兩個以上國家，其基礎建設或經濟成長等面向的發展狀況等等。最後，蘇昱璇助理教授向與會者展示 World Bank 網站的頁面，並一步一步的說明資料取得的流程。

王奕婷副教授分享的資料庫是「V-Dem」以及「V-Party」，並說明 V-Dem 計畫的動機，源自於民主概念的複雜性以及過往其他資料庫測量方式的侷限，並強調 V-Dem 的目的，是擴充而非取代既有的民主測量方式。接著，王奕婷副教授表示 V-Dem 的研究方法是專家調查，透過將大概念區分成具體而細小的問題，捕捉全球各地 200 個以上的國家，在不同年份的民主表現。另外，王奕婷副教授提到由於民主是無法直接被觀察到的概念，因此在運用專家調查測量時，可能存在專家之間標準不一致、對他國狀況不熟悉等問題。而 V-Dem 則試圖透過邀請在地專家以及運用試題反應模型，建構兼顧標準一致與在地知識的民主測量指標。

蘇彥斌教授談論的資料庫是「群眾動員研究計畫」，並表示該資料庫的主要目的是瞭解民眾對政府的抗議事件，是一筆分析單位為事件的橫斷面資料，包含 162 個國家自 1990 年到 2020 年的反政府抗議事件，總共有 17,415 個觀察值。蘇彥斌教

授接著說明資料庫的研究方法，是研究團隊在四個主要媒體，輸入「protest」等關鍵字，搜尋相關的抗議事件進行編碼，並指出該類研究方法往往存在未被報導的事件會被忽略的問題。最後，蘇彥斌教授提出幾個可能的研究設計、提供其他的抗議事件資料庫，以幫助研究者掌握此類型資料庫的使用方式。



第二場次的主題是「國家層次資料（II）」，由中央研究院社會學研究所的鄭雁馨副研究員，以及中研院法律學研究所法實證研究中心的主任張永健研究員擔任講者。鄭雁馨副研究員分享的主題圍繞在「人口與家庭變遷」，並分別介紹總體層級的「Human Fertility Database」，以及個體層級的「World Value Survey」與「Generations and Gender Survey」。首先，鄭雁馨副研究員表示 Human Fertility Database 至今已收集 33 個國家的人口生育資料，其中包括台灣，而運用該筆資料發表的出版品已經累積超過 600 篇，並舉例有研究論文以 Human Fertility Database 推估已開發低生育率國家的完成生育率。進到個體層級的資料庫，鄭雁

馨副研究員提到 World Value Survey 是一項橫跨 40 年的大型國際調查，目前為止有 60 約有國家參與其中，並以她本人的研究為例，說明如何運用 World Value Survey 探究儒家與非儒家圈對同性戀者的態度差異。最後，鄭雁馨副研究員介紹發跡於歐洲的 Generations and Gender Survey，表示該資料庫的目的是收集有關生命歷程與家庭動態的調查資料，並提到台灣正準備參與該項計畫，希望有更多有興趣的研究者加入研究團隊。

餐敘前的最後一場分享，張永健研究員介紹他所用過的三個法實證資料庫，以及由他個人完成的物權法資料庫。張永健研究員表示物權法資料庫的建置非常困難，他必須閱讀許多國家的法律，以確保資料庫中的每一個值都有所本。經過十年的努力，該資料庫已經收集世界各地 156 個法域的法律資料，並且有超過 200 個變數。而論及如何運用法實證資料進行研究，張永健研究員提醒研究者需要非常謹慎，因為壞的研究會讓世界變得更壞。他以 1998 年的一篇文章為例，該篇文章試圖顯示普通法系相較大陸法系更有助於國家發展，並影響世界銀行的放款政策，要求開發中國家在申請貸款時，需要修改國內的法律體系。然而，該篇研究沒有提供編碼簿，編碼基礎難以辨識，研究結果令人質疑。張永健研究員接著分享許多以法實證資料完成的研究成果，例如以 The Comparative Constitutional Compliance Database 的資料進行分析，發現已開發國家的憲法實踐較接近憲法規範，但開發中國家的憲法實踐與憲法規範就較為分離。

餐敘過後，工作坊的進行來到第三場次，該場次介紹的資料庫聚焦於個人層次及公司層次。第六位主講人為中央研究院政治學研究所吳文欽副研究員，吳文欽副研究員介紹的資料庫為選舉體系的比較研究（Comparative Study of Electoral Systems, CSES）以及全球民主動態調查（Global Barometer Surveys, GBS），前者關注的核心內容如投票行為、政治態度、對於政體的評估等，我國的台灣選舉與民主化調查（TEDS）於 1986 年加入 CSES，從 Teds 官方釋出的幾波面訪或電訪案中，便可以發現有幾道題目的調查註記了亦是屬於 CSES 的內容，可供研究者

作為跨國民意調查比較的參照。另外一個資料庫 GBS 則是由五個區域的調查研究所構成，分別為 Afrobarometer、Arab Barometer、Eurasia Barometer、Asian Barometer、Latino Barometer，該資料庫關懷的內容為民眾對於民主的支持程度、認為民主有哪些構成要素、對於民主體制的認知與理解等等。台灣所屬的 Asian Barometer Survey (ABS) 於 2000 年成立，目前正在進行第六波調查，值得一提的是，該調查也討論到民眾對於中美關係的評估與看法。



第七位主講人為中央研究院歐美研究所黃敏雄研究員，黃敏雄研究員介紹的資料庫為國際學生能力評量計畫 (Programme for International Student Assessment, PISA)，該資料庫主要的調查對象為各國即將或剛完成義務教育的十五歲學生，自 2000 年起每三年針對學生在閱讀、數學、科學等方面進行調查，研究目的首重學生在學習過程中的「素養」培育，探討學生能否了解現今社會生活所需的知識及技能，

以及能否將學校所學運用於現實生活中。除了資料庫的介紹外，黃敏雄研究員也向我們展示了他運用該資料庫進行的研究，主要討論不同社經背景的學生，在透過現行升學制度的成績分發下，會不會有更高的可能聚集在同一間學校中。該研究發現，不論是在過去的聯考或基測時期，甚至是現今採用的會考，都無法提升「不分家世，同讀一校」的機會。另一方面，該研究也發現，學生在學科素養上的差異主要存在於校際之間，隨著我國在考試制度上進行調整，採行繁星入學、會考等第制、就近入學等改革措施，不同社經背景的學生有機會就讀同一所學校的可能性，在2014年之後也顯著提升。

第八位主講人為臺灣大學會計學系劉心才副教授，劉心才副教授為我們介紹的資料庫為路孚特（Refinitiv Datastream），該資料庫被稱為英國版彭博（Bloomberg），在全球金融市場的數據中扮演重要角色。該資料庫的主要內容如成交量、上市公司的財務報表揭露、基金價格與流向等，內容包羅萬象，從價格數據、財報數據、總體經濟數據到分析師數據等等，能夠提供研究者豐富的金融市場資訊。除了資料庫的介紹外，劉心才副教授也向我們介紹了如何使用 Excel 操作該資料庫。最後，劉心才副教授為大家介紹應用該資料庫進行的研究，例如公司的資訊透明度在一國財稅差異程度的調節下，將會影響該公司的銀行聯貸案之合約內容與利率。

工作坊第四場次的內容主軸為資料庫的維護與處理，第九位登場的主講人為中央研究院社會學研究所的林宗弘研究員，主講的內容為資料庫整併與管理。林宗弘研究員認為增加資料庫中的變數與時間尺度，有助於發現新的理論，同時也能夠檢視理論的穩健性與可運用之時間範圍，另一方面，在因果推論上，也有利於找尋工具變數及進行反事實推論。林宗弘研究員建置的國際資料庫涵蓋1980迄今的215個國家，資料來源大約40個資料庫，總共16000個變數，主題涵蓋民主體制、性別政治、經濟、能源、氣候、災難與衝突以及公共衛生等。林宗弘研究員亦提及不同資料庫在整併上可能會遇到的問題，例如系統性遺漏值、更新頻率不同、國家的分裂與新生等等，因此在資料庫整併上往往需要進行人為的加工處理，即透過插補的方

式極大化研究者能夠使用的資訊，進一步而言，資料插補也有幾項大原則，如從官方或公開資料進行資料填補、依變數無污染原則（完全不插補）、控制變數補完原則（減少多個資料庫的多重流失）、貝氏插補優於線性插補等等。

最後一位主講人為中正大學心理學系蘇雅蕙副教授，主講的內容為多層次模型與試題反應理論。多層次模型主要用於巢狀資料結構，並預設了同一個組內的性質是類似的，例如分析不同學校中的不同班級內的學生，該分析單位便同時涉及了學校、班級、學生，而多層次模型讓研究者得以將不同分析單位相關的變數都放入模型中討論，好處為能夠避免變數間錯誤關係的推論。接著蘇雅蕙副教授介紹了試題反應理論（Item Response Theory, IRT），透過項目特徵曲線（Item Characteristic Curves, ICC）能夠在題目設計上反應，一個人具有能力回答該題目並能夠正確作答的機率曲線，該理論在應用上的長處為能夠客觀地排除題目的難易度與個體的智商差異，因此進行題目的難易度估計上能夠排除樣本的影響，在這樣的優勢下，使得項目反映理論目前在學界的應用，得以取代古典測驗理論。