

資料文化新趨向 統計調查 新視野

社會數位轉型造成大量資料融入於日常生活中，資料文化正悄然變化，各種創新思維及技術方法，促進多源資料範式發展，面對資料生態系統的變革，調查工作應有與時俱進之統計意識，積極擴展新視野，以開創統計調查新局。

羅國華（行政院主計總處國勢普查處研究委員）

壹、前言

長期以來統計調查廣泛提供各界需用經社指標，這些經設計而蒐集的資料，有統計理論的驗證與理解，且以資料為核心持續發展所形成的資料文化亦廣為人熟悉。邁入 21 世紀，社會數位轉型帶動新型態的經社活動，產生大量資料且樣態多元複雜，資料生成及應用範式受到衝擊，政府統計所遵循的規範及標準受到檢驗。面對資料變革新趨向，統計調

查面臨嚴峻挑戰，應有與時俱進之思維，擴展新視野以順應變化。

貳、變化世界中的資料文化

政府統計的編製具有多方面概念，深受資料性質影響，其來源分為主要來源（如統計調查）及次要來源（如行政登記、新數據）。統計調查是政府統計資料不可或缺的主要來源；行政作業資訊化促進資料的連結運用，而減輕受訪者負

擔並擷節成本。近年資通訊科技蓬勃發展，各種電子商務交易、行動裝置、感測設施、衛星影像、物聯網、網路行為與社群網絡等新數據，持續源自公私有機構，具複雜且迥異之性質，帶來蒐集方式、處理技術、統計方法、品質標準、隱私安全及人員素養等方面的新課題，造成資料文化悄然變化，因此，國際統計組織積極進行相關的研究及推展，各國統計機構愈益重視新數據的開源及應用，以掌握契機因應變革。

一、發展多源資料 促進統計創新

統計機構致力於資料發展，將行政資料應用在編製統計、替代或補充問項、抽樣底冊、結構權數、驗證、檢誤及插補等，精進統計調查成效顯著。2006年歐盟統計局提出「調查與行政資料的整合（ISAD）」計畫，進行相關研究並共享經驗，而成爲歐盟國家重要的統計生產方式。資料整合能豐富資料變量及細緻度，擴增廣度與深度，促進統計創新之契機。例如美國普查局「雇主與家戶縱向動態

（LEHD）」計畫，整合聯邦及州政府多種勞動供需登記資料及普抽查資料，而創新產生地區季度勞動力指標（QWI）及轉換工作流動（J2J，附圖）等統計；行政院主計總處（以下簡稱主計總處）積極整合多種登記資料及普抽查資料，創編多元薪資統計及縣市別薪資統計等，擴增統計應用。

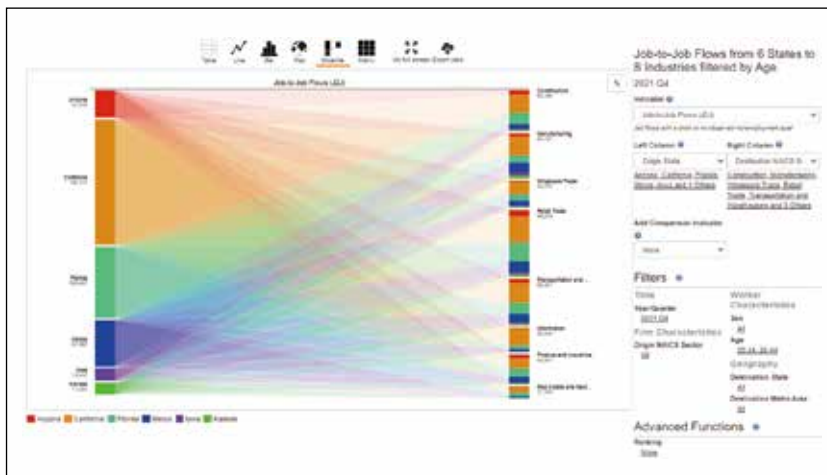
數位時代各種經社活動生成的數位足跡無處不在，具有實時、系統生成及相關的特性，使「資料」的含義及關係產生變化，如何從大量資料中提取有意義的訊息以創新統計，成爲統計發展的新趨向。例如澳

大利亞統計局分析航照/衛星影像識別新建物及其類型，自動化進行地址分類對應，以取代調查；加拿大統計局將衛星影像與作物保險及農業氣候數據結合，進行作物分類及產量估算；進行CPI及零售商品計畫，以掃描數據替代部分調查數據；愛沙尼亞利用行動數據於旅遊統計，產生更及時且精確的空間及時間估計。

二、善用新技術 擴增資料應用

調查資料是基於統計方法，行政資料是基於管理需求，產生過程均受到權責單位的規範及管控。新數據是基於不同目的產生，資料結構及內涵品質差異大，傳統統計技術適用性受到侷限，致各種新技術應運而生，涵蓋跨領域的知識。面對資料環境變革，聯合國歐洲經濟委員會（UNECE）政府統計現代化小組，積極合作進行相關技術的研究及推展，2017年提出資料整合指引，提供整合方法及框架，衡量品質維度及實施建議等；2019年開

附圖 轉換工作流動（Job-to-Job Flow）



資料來源：美國普查局網頁查詢系統。

專題

啓機器學習相關研究，陸續進行 19 項試驗，另提出統計算法的品質框架作為使用指導，透過經驗共享逐漸成為最具有發展潛力的新技術。

運用鍵值 (key) 連結是資料整合的重要方式，除橫斷面 (cross-sectional) 鏈結應用，若行政資料個體按時序串接，可產製縱貫面 (longitudinal) 的追蹤資料集，擴增應用範疇。例如美國「人口普查縱向基礎建置計畫 (CLIP)」，係整合歷次普查、社區調查 (ACS) 及當前人口調查 (CPS) 等為核心資料，鏈結行政資料或其他外部資料，提供豐富的世代連結及社會長期發展變化之縱向資料。機器學習具有處理及分析多類型且大量複雜資料的能力，包括圖像、影音、文本等非結構化資料，經由資料訓練中自動學習經驗，而不斷探索資料之規律及變動性，再據以建立預測模型，以對未知資料進行預測，UNECE 應用於資料的編碼、分類、檢誤、插補及圖像分析等方面，各具成效。其他技術如網路爬梳用於

自動存取網頁內容、文字探勘是從非結構化的文字資料中提取出有意義的資訊等，都是探索新數據的新工具。

三、保護資料隱私 建立合作關係

保護資料隱私是統計人員作業的基石，統計法第 19 條規定統計調查取得之個別資料應予保密，除供統計目的之用外，不得作為其他用途。統計調查利用統計披露控制 (SDC) 方法，對資料進行概化、抑制或擾動值的措施，以降低個別資料重新識別的風險。現有更多來自其他公私有機構之次要來源資料，不僅細緻度增加，且經連結擴增屬性，亦會增加資料敏感度，且其各有資料隱私保護的特定法規及限制，使保護需求益顯複雜與重要。聯合國統計委員會 (UNSC) 設立隱私增強技術任務組 (Privacy-Enhancing Technologies Task Team)，建立網路協作平台，各國專家共同探索相關技術、原則及標準，2023 年發布「隱私保護技術手

冊」，提出安全多方計算、同態加密、差分隱私、分佈式學習、零知識證明、合成數據及可信執行環境等保密技術，旨在安全處理及共享敏感資料，並以多個試驗案例分別說明。

面對資料生態系統的變化，為消除資料使用障礙及保護隱私，並能長期生成值得信賴的資料，UNECE 及 UNSC 建議統計機構應加強與資料擁有者建立合作關係，以能討論及預見資料的變化。例如有些國家統計局 (如英國、荷蘭、瑞士) 創設資料科學中心，引進人才進行資料技術研究及整合推展，並與相關領域夥伴加強關係促進交流。對於統計機構使用私有機構之資料，聯合國建議基於社會責任，資料的提供應是自願且免費的，即資料本身不應計價，但可合理補償資料提取成本及預處理費用。

四、創新資料思維 加強資料素養

數位革命使資料成為用之不竭的新資產，融入於日常生活中，如何從大量資料中探勘

出有價值的訊息，是資料生成的新思維，應用上須對資料具有更深的洞察力，以能發現其中隱藏的規律及現象。不同於統計調查係針對問題而蒐集資料，並基於機率進行資料檢驗及推論的過程，其是由資料驅動而採行相應的技術及方法，創新的思維是資料利用的泉源。例如荷蘭統計局在資料發展上，鼓勵人員不拘形式的發想及共同的腦力激盪，以激發出有價值項目，再經由概念驗證及試作實驗確定效用，再予轉換為成熟的統計應用。

資料是原材料，如何轉化以創造價值，是統計機構努力的方向，傳統靜態、定時的資料蒐集方法，面臨動態、整合的資料生成挑戰，原有生產框架難以適應而須相應調整。因此，加強多樣態資料的探索能力，學習跨學科技術及方法，以及建立合作夥伴間的知識交流，提升人員資料素養，以因應變革。

參、資料變革中統計調查之精進發展

來自次要來源的大量資料提供從多維度視角觀察，助於產生新的見解及認識，統計調查可控制設計彈性及獲取更豐富變量，助於衡量個體的具體特徵，兩類間各有不可忽略之功用，具相輔相成之效果。資料變革促成多源資料發展，新技術從根本改變了資料的生成方式，造成統計調查範式的衝擊，擴充資料來源增加統計新指標，成為統計調查發展新趨向。

一、整合行政資料強化調查統計

資料整合具有時效、成本、頻率、細緻度及涵蓋面等多方面效益，主要困難是整理來自於各機關的資料，在定義、分類、範圍及時期等的一致性，須先有充分的了解與評估。關於整合資料的品質衡量，可利用相關的調查資料進行機率錨定（probability anchor）。在整合概念帶動下，主計總處積極應用行政資料，產製更多樣的統計結果，替代或簡化統計調查。例如連結戶籍、入出國、

勞保、健保等資料相互勾稽，並用人力資源調查資料錨定機率，創編國人赴海外工作人數統計；利用戶籍檔、出生、結婚登記、勞保、健保（農保）、人口普查及工商普查等資料，整合創建女性職涯縱向動態資料。行政資料應用價值已廣受重視，許多國家（如荷蘭、韓國、加拿大等）統計法授權統計局為統計目的需要可使用相關行政資料，我國統計法第 18 條亦有相同規定，因此在資料發展上應以行政資料應用為優先，再輔以統計調查補充不足，相輔相成豐富統計結果。

二、運用新技術創建預測模型

機器學習是新數據生成最常用的技術，其演算法複雜常被認為是難以理解的黑匣子，透明度受到質疑，使用上應加強資料可解釋性的說明。統計生產重視估計品質及變量間關係，機器學習則注重預測的準確性，對統計機構而言是頗具挑戰性的新技術。主計總處整合跨部會公務登記資料，

專題

藉由不同變數組合之拔靴取樣 (bootstrap sampling)，再以機器學習方法進行資料訓練，進而建立常住人口預測模型，提供常川編製區域別常住人口統計。機器學習的資料生成範式迥異於傳統的推理範式，目前仍以試驗性質居多，惟在多種類型資料上具有很大潛力，UNECE 項目報告結論認為機器學習領域正快速發展，新方法不斷出現，應持續關注跟上步伐。

三、推動數位普查提升調查效能

三大普查動員人力均達萬餘人，以及民衆對隱私權的重視，致調查環境日益艱困，促使數位化進程加速，包括優化行政作業系統，使人員、工作配置、經費、考核等作業更有效率，並跨接網填系統，加強共通資料利用；建置網填系統，方便受訪者填報，具線上即時檢誤能力，並依各級普查人員權責，提供相應之資料審核及作業管理功能，可及時掌握進度與品質；提供行動裝置上網

填表，導入驗證技術，建立傳輸加密安全機制，強化隱私保護；推動線上會議及訓練，編製數位教材；開發輔助審核系統，進行線上傳輸及審核作業。普查期間適逢 COVID-19 疫情肆虐，惟因各項作業多已數位化，且網路填報宣導奏效，因此訓練、填表及審核等作業均可網上執行，使疫情影響大幅降低，而能確實完成數位普查。網路填報不受時空限制，可避免接觸訪員而強化隱私，亦能減輕蒐集作業負擔，增進調查效率，已成為各國積極推展之資料蒐集方式。

肆、結語

資料革命帶來資料多源及技術革新，整合方法擴增資料的應用價值，新數據改變資料的生成範式，資料文化正悄然變化，衝擊既有的運作模式，歐盟統計局指出統計系統應以資料擴充形式發揮各自價值。面對變革，展望統計調查應跳出固有思維模式，關注多源資料生成效益，加強合作夥伴關係及知識交流，提升人員資料

素養，並汲取各國經驗及做法，善用新技術及方法，擴大資料生產格局以增進應用價值，開創統計調查新局。

參考文獻

1. 羅國華 (2021)，精進技術發展為資料創新鋪平道路，主計月刊，789 期。
2. United Nations (2023), *United Nations Guide on Privacy-Enhancing Technologies for Official Statistics*, United Nations Committee of Experts on Big Data and Data Science for Official Statistics.
3. Sofie De Broe, Peter Struijs, Piet Daas, etc. (2020), Update the paradigm of official statistical : New quality criteria for integrate new data and methods in official statistics, *Statistics Netherlands*.
4. Machine Learning Team (2018), The use of machine learning in official statistics, *UNECE HLG-MOS Machine Learning Project*.
5. High-Level Group for the Modernisation of Official Statistics (2017), *A Guide to Data Integration for Official Statistics*, *UNECE*.
6. United Nations Global Working Group on Big Data (2016), *Recommendations for access to data from private organizations for Official Statistics*, *UNGWG*. ❖