

厚資料與意義探勘專刊導論

Introduction of Special Issue on Thick Data and the Meaning Mining Approach

劉正山

國立中山大學政治學研究所教授兼所長

本刊的發行時間，距離世上第一支 iPhone 上市（2007 年 6 月 29 日）剛好滿 12 年。隨著行動裝置、Web 2.0 社群網路的快速發展及資料量快速累積，這可以說是個資料和技術急速發展，令許多文獻無法解釋新興現象與問題的時代。這也是個學界面對世界轉變、學界新人興趣出現轉向的時代。近年，大數據（big data）的概念風行，帶動眾多研究領域（如心理學、管理學、傳播學、語言學）的轉型。我們已經見到許研究者從規模龐大的數據中爬梳規律及新知。他們為 1970 年代就有的探索式的資料途徑拍去了陳年積灰，而這股「資料導向」（data-driven）與「方法導向」（method-driven）研究的風潮讓本來就很強勁的「因果推論」（causal inference）研究風起雲湧，愈來愈多研究者使用更多、更新的資料與方法進行實驗與探索。雖然這個風潮還熱不到「跟不上就完蛋」的局面，但是對於「數據可取用性增加」不感興趣的學者，或是對於已經能夠看見資料分析價值但不可能為此放下原本研究取徑、重新學習程式語言的學者來說，現在十分強調資料（含大數據）及方法的風潮，可能已令不少學者感受到缺乏話語權的處境。

本刊的意義在於展現「非科技人」如何在這個資料時代重新掌握知識話語權的視角及想像。這些作品的對話的對象，其實並不是「大數

據」領域，而是研究者自己。本刊五位作者所共同勾勒的「厚資料」的視角及做法，將有助於讀者看見「理解並發掘資料背後意義」的這件事對知識建構的價值。

厚資料（thick data）這個名詞大約在 2013 到 2014 年間被創造出來。先在網路上流傳，後來出現在管理學的評論及期刊之中。一開始，這個詞的意思是強調「質化」方法的知識建構，多是從人類學的視角出發。但這並不新。其實，「厚」的核心內涵很早就人類學中被運用，原稱叫作厚實描述（thick description），因此，現在使用「厚資料」一詞者，不少是從「厚描述」或「厚敘事」（thick descriptions）的人類學民族誌研究方法（ethnographic methodology）傳統來使用這個詞。可惜的是，若只是這樣從「研究方法」的角度來看這個詞，那麼「大」數據與「厚」資料，可能只是幾個世代下來「量化」與「質化」之爭的舊酒新瓶。

我們需要更持平地、用更開闊的角度來看待這個詞。本刊所收錄的四篇文章，是 2017 年科技部人文社會科學研究中心「厚資料知識建構」學術研究群的部份成果。在那一年中，我擔任主持人，舉辦了七場會談，讓近二十位國內政治、心理、社會、公共行政、管理等領域學者從真切的論辯看見彼此的知識論立場，讓彼此在對話中逐漸看見「厚資料」的意義。一年下來，大家都同意，所謂的資料，其內涵絕不限於數量型的資料（numerical data）。但另一個大家比較不情願去同意卻必須面對的事實是，學者不見得是真正在意資料中挖掘意義這件事。我們看見，從事數量資料分析者，很可能只管「借」別人的理論來進行假設檢證，期勉自己用期刊接受的方法及格式來完成論文（就好），無心進一步深化挖掘分析結果的意義。我們也看到，從事質性研究的學者，也很可能只強調自己個案的敘事而拒絕尋找可供客觀對話的數據來完成整個故事。這是很難得的發現。這一年下來，我們透過介紹自己的知識論及方法論，看見彼此的限制，這些真誠的對話鍛造了這本特刊，甚至延

續至 2018 年的中國政治學會及 2019 年的台灣政治學會的「厚資料專題論壇」，以及其他正在投稿中的作品。這一兩年下來的心得，讓參與對話的學者有機會看見：數據的大小並不是重點，用不用大數據也不會是新浪後浪的壓力源。真正重要的，是資料的收集者與使用者並不是站在意義探勘的立場看待自己能做出來的貢獻。學群的成員從年頭談到了年尾，最後很驚喜的發現，幾個重要研究方法背後其實有些相近的，對於「探索 + 描述 + 詮釋」的期待。在這個厚資料方法論的視野下，我們沒有必要再去緊抱單一的質性方法或量化方法；因為只要能產出更深入瞭解現象背後意義的方法就會該被視為好方法。從知識論的角度出發，厚資料會是一種求知的途徑，可以被實證主義、科學實存主義、實用主義、詮釋主義所共享。換言之，厚資料與大數據本來不該是個方法上質量之爭的兩方，而是對意義解讀的期待能否落實到不同型態資料的分析之上的一種立場與態度。

若我們能從這個脈絡來思考，那麼就能發現，「厚資料」就不只質性資料的新代稱。精準的說，厚資料是個「對資料進行分析之後採取發掘其意義」的途徑（Data-based Meaning Mining Approach, DAMM）的代稱。我猶記得學群成員何宗武教授曾主動討論到 Meaning Mining 與 Meaning Netting 的差異。對我們來說，這兩者不只是英文縮寫是不是文雅易記的差異而已（我承認，他所提出的後者其實更容易辨識易記）。我選用的 Meaning Mining 一詞承載著一如 data mining「向深處挖掘意義」的期待；而他建議的 Meaning Netting 則強調了意義之間橫向張網般的連結。無論日後學界選用的名詞為何，都不影響厚資料途徑對於一個掌握客觀資料（含大數據）的研究者的責任：要能為讀者揭露資料背後更深厚意義。對一個採取厚資料途徑的詮釋主義研究者來說，即便再怎麼強調主觀詮釋的重要性，他/她會把客觀的資料納入論述，並致力於對資料進行正確的分析。

劉致賢的「初探厚資料與中國大陸政治經濟現象：以國有企業部門

為例」更是率先展示了如何以厚資料途徑，來面對中國大陸國有企業部門相關的經濟活動數據失真的問題。非常難得的是，劉致賢教授將他長年投入中國大陸研究與觀察的心得，連結到厚資料途徑，清楚的勾勒與說明這個途徑對區域研究的助益。劉教授強調，意義的產製來自於人，也就是行動者。取用厚資料途徑的研究者，無疑會是個資料意義的開發者，但我們不能忽略資料來源或行動者（例如中央企業、地方國企、金融機構、中央政府與各級地方政府本身等）也是意義的創造者。因此，若要掌握分析結果的意義，就需要瞭解資料的來源、脈絡及行動者的動機。此文清楚的展示了劉教授如何將文獻中的資訊匯整為理解脈絡的框架及機制，進行辨識出造成檯面上新聞矛盾訊息背後的行動者其實是不同的群體。從這裡開始，我們就能看見即使是小數據或是不真實的資訊，背後真正能幫助我們瞭解中國大陸國企圈中的政治，以及日後從不同集團行動者發出資訊時需要注意的地方。

鄭安授的「福島事件後台灣核能議題的媒體再現」這篇則是展現了傳播學者如何透過厚資料途徑的觀點，將描述性的資料分析增加意義。這一篇研究使用了內容分析法，觀察2011年福島核災之後的十四個月內國內三家報社對此事件的報導，由此找出框架（新聞播報者的視角）以及主要的行動者（消息來源）。這個做法與上一篇劉致賢教授的建議相當一致：鄭教授在發掘出行動者本身的立場與價值之後，進行「新聞媒體報導所呈現趨勢背後的意涵解釋」。這篇文章透露出一個訊息：在厚資料途徑在乎的是資料背後的意義。只要能抽取出意義，數據的大與小，以及技術本身並非重點。這篇研究透過小數據萃取出來的樣貌，更新了我們對事件背後的理解。

邵軒磊與吳國清教授的「法律資料分析與文字探勘：跨境毒品流動要素與結構研究」則同樣是使用文本進行分析。特別的是，他們以法院判決書作為礦脈進行意義探勘。這一篇研究的資料量更大，但關懷仍然維持在從資料中萃取意義。兩位作者如何形成問題意識，以及為何走向

選用文字探勘的動機，非常值得讀者的重視與參考。此外，這一篇是本刊中最接近結合「厚資料途徑」與「大數據方法/工具」的實例。他們嘗試從文字資料中描繪跨境毒品的流動結構，並進一步找出相關的因素。作者透過尋找欄位之間以及關鍵字之間的關係，讓一般用於描述之用的大數據資料，展現了「協助研究者發掘問題」的潛力。雖然我個人將厚資料視為一種方法論，但作者「將厚資料看作一種足以萃取出意義的資料型態的總稱」的視角，也相當值得參考。

彭睿仁的「聯邦體制與選舉制度對不同層級選舉間選民政黨偏好轉移之影響」：

德國 2013 年至 2017 年聯邦眾議院及巴登—符騰堡邦議會選舉之實證研究」這一篇則是個從實證主義的知識論體系中取用厚資料途徑的嘗試。彭教授在理論檢證的典範中原本可以直接使用個體資料來分析選民意向的做法之外，進一步取用了集體資料，以及額外的分析方法來觀察個人層次的投票行為。這個做法為實證主義知識論的世界擴展了更多的趣味及發掘意義的空間。另外值得注意的有兩點：(1) 彭教授示範了取用厚資料途徑的實證主義者，縱使將假設檢定通過與否設為首要目標，還願意「多說幾句」，認真道出這些資料及分析結果背後的意義；(2) 他直接公佈資料及分析的原始碼，讓資料分析過程透明、可再製。這將讓知識社群得以藉以參與未來同一個專案的意義建構工作。這個做法雖然尚未是學界共識或是學刊出版的要求，卻已是頂尖國際期刊設立的標準。這篇文章確實地展現了實證主義者如何取用厚資料途徑來展現研究成果，非常值得實證研究社群參考。

整體來說，本刊的四篇文章帶著大家從背景到論述，再到實例操作，先理論後方法、先宏觀後微觀。四篇之間雖然不是直接在主題上對話，但讀者會發現它們背後共享的厚資料方法論。這正是 2017 年厚資料學群學者的共同感受到的趣味。最後，我想強調，「路是人走出來的」，因此為路命名並不是重點。踏過棧道的開路人更不需要知道路

名。無論用的是「厚資料」或「DAMM / DAMN」，它終究只是個方便於溝通及傳遞概念的詞彙。我希望它是跨越不同知識論體系的方法論立場，為所有學者共享。因此厚資料途徑不會專屬於特定知識論體系，反而足以為目前的知識論體系所接受，成為各個知識論體系中，另一個知識生產途徑的選擇。若這個學群的努力與用心，以及本刊作者努力的心得，有助於學界持續溝通和踏出步伐，持續以厚資料的精神發表優秀的研究成果，就足已令我等慶幸。2007年起全世界才一起使用了智慧型手機；2016年我們也才與世界同步對大數據風潮進行了反思。現在我們有了這本厚資料專刊。我們現在的省思及行動，並未落後世界上其他優秀的學群太多。我們已鳴槍起跑，用新的途徑看待自己正在做的研究，或是讓自己踏上這條路徑來規畫未來要做的研究。在各自的專業領域中，「把問題做深」的努力，都會讓我們的研究產生獨特的光亮。這是本刊作者群以及2017年科技部人文社會科學研究中心「厚資料知識建構」學術研究群對於國內學界的貢獻與價值。