

政治學中建立厚資料傳統的契機與挑戰

劉正山

中山大學政治學研究所 教授

本文發表於 2016 中國政治學會年會

2016.11.25-26@東華大學

摘要

社群網路普及和大數據 (big data) 的思潮帶動了政府與企業對於資料分析的需求。即使大數據的發展正旺，對於大數據的限制及反思也正在企業管理、社會科學中興起。本文所倡議的厚資料 (thick data) 途徑是由資料中挖掘出意義，而不只去探索或描述趨勢。有別於傳統的用質化量化二分法來看待現今的這個大數據與厚資料風潮，也就是單純的把大數據等同於量化分析、或是把厚資料當作是質性方法的代名詞，本文將從知識論的論述出發，將大數據必需融合厚資料觀點，以及厚資料不必只是人類學的口號的思維系統性的呈現出來。隨著大數據分析的平民化及大數據在分析端終究要成為小數據的技術及實務面趨勢的浮現，政治學經驗研究的傳統，正視厚資料所代表的「人腦開發意義」的思維將有助於學門與企業及政府治理接軌。本次發表將從知識論及方法論呈現「厚資料途徑」的意義，並輔以網路調查平台 smilepoll.tw 的運作經驗作為例子，談談網路調查平台如何能夠成為訪談及收集定群追蹤資料 (同一群人不同時間受測) 的工具，同時成為厚資料方法論下所需要的資料收集工具。

關鍵詞：大數據、厚資料、知識論、方法論、網路調查

一、前言

(一) 大數據發展思潮與知識建構的反思

近 10 年可以說是個資料和技術引領求知的時代。隨著社群網路的快速發展及資料量快速累積，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) 掀起了「資料導向」(data-driven) 與「方法導向」(method-driven) 的研究風潮，研究者紛紛開始從規模龐大的資料中探勘現象的發展趨勢。

資料導向的研究成為顯學之際，我們也見到伴隨的反思風潮。強調規模龐大、牽涉廣而繁雜的資訊分析技術，容易讓學界容易減弱對資料意義發掘的注意力。若是將人文素材的數位化當作是研究課題，或是只著眼於技術應用層面，忽略知識論、方法論的知識建構層面，那麼人文社會科學的意義反而更加邊緣化 (項潔, 2011)。

意義是社會科學中非常重要的知識建構基礎，且這些知識基礎涉及了群眾及組織的情緒、個性、以及行為和偏好，它們無一不是當今社會所需要的知識面向。因此，社會以及社會科學界對於決策者及公民偏好的形成，以及行為及決策的原因 (包含觀念、信念，以及這些行為偏好、持續或改變的相關知識)，有強大的需求 (Madsbjerg & Rasmussen, 2014; Uprichard, 2015)。所以，若要深化民主，並且不要讓決策者只是跟在資料後頭決策，我們還需能掌握資料所提供的意義、價值及述事，來呈現公民的價值取向。

大數據在社會科學的的功用是提供「更豐富版本的關於人類行為及互動的資料」，也就是擴大現有的資料範圍，讓取得的資料來協助設計更精確的實驗、進行更好的母體群體之間的比對，並觀察到以前較難以觀察到的人類行為 (Monroe, et al., 2015)。換言之，大數據能協助實證主義者仰攻「因果推論」(causal inference) 這個更高的目標。相對「小」的資料 (例如訪談、民意調查資料等) 或許不具技術上的價值，但仍不失在社科上的價值 (Gray, Jennings, Farrall, & Hay, 2015)。對於實證主義的知識建構立場來看，小資料能夠協助產生假設 (Rasmussen & Hansen, 2015)。

不過，即使用這個角度來為「小數據」加值，也無法掩蓋目前大數據潮流所面對最大盲點，就是專注在挖掘資料及發展資料分析技術的同時，正視發掘嵌在資料中豐富意義與故事的價值，以及學習相關的研究方法。因此，這個大數據目前還貢獻不到的非實證主義知識建構，非常值得社科學者 (尤其是有相當方法論傳統的政治學者) 投入研究和論述。

(二) 現行非實證主義知識論體系的紛雜

實證主義的論述無疑是現今社會科學的主流，從理論建立、假設推演、資料檢證假設的知識論、方法論及研究方法發展十分成熟，近年更走向因果推論 (causality) 及實驗方法。不過，當前實證主義之外的知識建構體系出現紛雜，眾多研究方法在沒有知識論作基礎的狀況下各自為政的現象。

知識論體中，除了實證主義之外，尚有科學實存主義 (scientific realism)、實用主義

(pragmatism) 及詮釋主義 (Interpretativism) 等。科學實存主義與實證主義呼應，但更強調機制的重要；用實用主義強調行動與解決問題才是知識；詮釋主義則以反實證的角度強調人的問題的複雜性及不可化約性。因此，由實用主義及詮釋主義延伸出的方法論及研究方法不同於以假設檢證及邏輯推導為核心的實證主義及科學實存主義。

當前大數據的風潮所帶起的資料科學 (data science)，強調資料導向研究 (data-driven)，重視對資料的「觀察、探索與意義發掘」。這反襯出了目前社會科學界在追求理論導向研究 (theory-driven research) 時對於「觀察、探索與意義發掘」的冷落。data-driven 與 theory-driven 之間不但出現了技術上的新壁壘，甚至也出現了 data-driven 為貴的學門對以 theory-driven 為貴的學門之間相輕的現象。社會科學因此必需面對從研究議題乃至就業市場上被邊緣化的困境。

因此，非實證主義的各種研究方法，以及這些方法的價值需要被整合、討論及再認識。誠然，在諸多不同的知識論及方法論途徑中我們不可能面面俱到，甚或是天真的以調和論來追求某種知識論上的一統。最可行的思考方式，是在知識論與方法論之層次，找出能夠推動「觀察、探索與意義發掘」的方法論，並整合相關的方法和工具。

(三) 數位人文的發展趨勢

數位人文一詞起源於二戰之後由 Roberto Busa 神父啟動的「人文計算」(humanities computing) 研究，亦即以電腦輔助記錄經典的文本，以便進行更有系統的神學研究。到了 1980 年代隨著電腦技術的發展，Text Encoding Initiative 計畫便開始展現人文結合電腦的企圖心。就窄義來看，數位人文就是結合大量數據材料，運用資訊科技，來進行人文研究，目的是重新發掘出「前數位時代」中難以觀察的現象、無法想像的議題與無法進行的研究，其背後的價值 (項潔, 2011a)。就廣義來看，數位人文這個詞的定義已超過二十個。¹¹ 由近年出版的學術著作及國際學者作品可以看出，數位人文已超越了「數位典藏」，不再只是強調電腦與人文的合併概念而已，而是成了「利用資訊科技從事人文和社會研究」的代名詞 (Gardiner & Musto, 2015; Gold, 2012; Schreibman, Siemens, & Unsworth, 2016)。台灣自 2002 年起追上這股潮流，尤其自 2009 年起國立臺灣大學數位典藏研究發展中心舉辦了第一屆「數位典藏與數位人文」國際研討會、推出一系列數位人文研究叢書討論數位人文在史學及文學的應用之後 (項潔, 2011a; 2011b; 2011c; 2014) 帶來了重新聚焦的效果。數位人文的理念強調 1. 產生文本礦脈；2. 將人文與社科學門中重視意義發掘的傳統重新標舉出來；3. 結合資訊科技，讓過往可能流於過度主觀的詮釋 (例如，挑選關鍵詞本身就是一件十分主觀的事)，找到一個客觀的基礎，讓眾多文史社科的研究者能在的相對主觀的描述與詮釋上，建構相對客觀的知識。

這個數位人文理念與上述對於社會科學資料的「觀察、探索與意義發掘」關懷是一致的。這個關懷中，與社科最相關的研究方法是文字探勘方法 (text mining)。這個方法利用「詞頻分析法」及「切詞」技術，以統計方式處理一個或多個或多組「關鍵

1

http://www.artsrn.ualberta.ca/taporwiki/index.php/How_do_you_define_Humanities_Computing/_Digital_Humanities%3F

字」，作為資料探勘 (data mining) 的資料集。細言之，這個研究方法包括兩個環節：首先，研究者計算「關鍵詞」出現頻率來呈現某種現象的趨勢，再結合詮釋來呈現歷史事件或概念形塑和演變過程。接下來，研究者找出關鍵詞之間的關聯或叢集，進行更深入的現象背後的意義發掘。舉例來說，在史料文本中檢索「朱一貴」三字，除了可以發現朱一貴事件直接相關的人物外也可以看現有相當數量關於林爽文事件的人物。讓兩者之間的潛在關聯浮現，值得加入更多解釋與詮釋 (項潔、涂豐恩, 2011)。目前，世界各主要從事數位人文的文字探勘工程團隊之中最受重視的，應屬哈佛大學政府系的 Gary King 教授的團隊。Gary King 團隊目前最受重視是研究成果，是利用文字探勘方法分析社群媒體網上留言，發現中國政府在網路上輿論監控時首要的並不一定是政治敏感字眼，而是可能引發集體行動的字眼 (King, Pan, & Roberts, 2013)。這個知識建構的方式，與統計學中一個重要傳統——探索式資料分析 (exploratory data analysis, EDA) 十分相似，只是研究工具不同。若是不同領域學者之間確實依循存在著一個共同的知識發展信念 (知識論)，那麼這個研究途徑就值得被進一步勾勒出來。

二、厚資料：是一種方法論 (途徑)，而不是一種資料

從資料中挖掘意義其實並不算新的概念。它的核心內涵很早就有人類學中被運用，這個方法叫作厚實描述 (thick description)，由 20 世紀美國人類學家 Clifford Geertz (1973) 提出。從人類學到社會科學有很多學者都在從事人類行為及文化的觀察工作，他們擁有對脈絡掌握及述事的能力，所收集到的、具有豐富意義和故事的資料其實就是厚資料最原始的意思。「厚描述」或「厚敘事」 (thick descriptions) 的概念，是今日人類學民族誌研究方法 (ethnographic methodology) 的核心價值。

厚資料 (thick data) 這個名詞則是大約在 2013 到 2014 年間被創造出來，一開始是針對大數據強調的 data 一詞，從人類學的立場來強調能夠持續被開發出新義及故事的資料的重要性。就從這個「探索+描述+詮釋」的目標來看，這個名詞已超越人類學，反映的是個方法論的立場。畢竟從人文到社科，資料的型態多樣 (例如文本、圖片、聲音、影片等)，不是只有人類學的資料才是 thick data。所以 data 一詞無法且不宜被窄化為 (大數據概念下) 只能供機器讀取和分析的資料，或是純然 (人類學觀點下) 田野調查的資料。

thick data 一詞一起始看起來像是 big data 的對照，但是仔細推敲後會發現，大數據與厚資料它們並不是相反詞，也不是單純的互補。Thick data 被賦予了知識論上「探索與詮釋」的期待。我們可以看到厚資料不必然需要成為大數據；而當大數據能被開發出豐富意義的時候，大數據也會是厚資料。這個「未來的研究必需平衡大數據與厚數據」的觀點目前已在管理學界及實務界流傳和討論 (Madsbjerg & Rasmussen, 2014; Tencer, 2014)。

至於厚資料裡所強調的詮釋，很容易令人聯想到詮釋學派的知識論立場。詮釋學派關心人們的社會行為所賦予的意義，從而對文本 (text) 和行動加以解釋或發現某些事件的關聯，得出有關於社會的詮釋。對詮釋學派來說，由於意義依附在特定社會及時空上，因此進行厚描或敘事都是暫時、局部的主觀詮釋 (Marsh & Stoker, 2002, pp. 27-

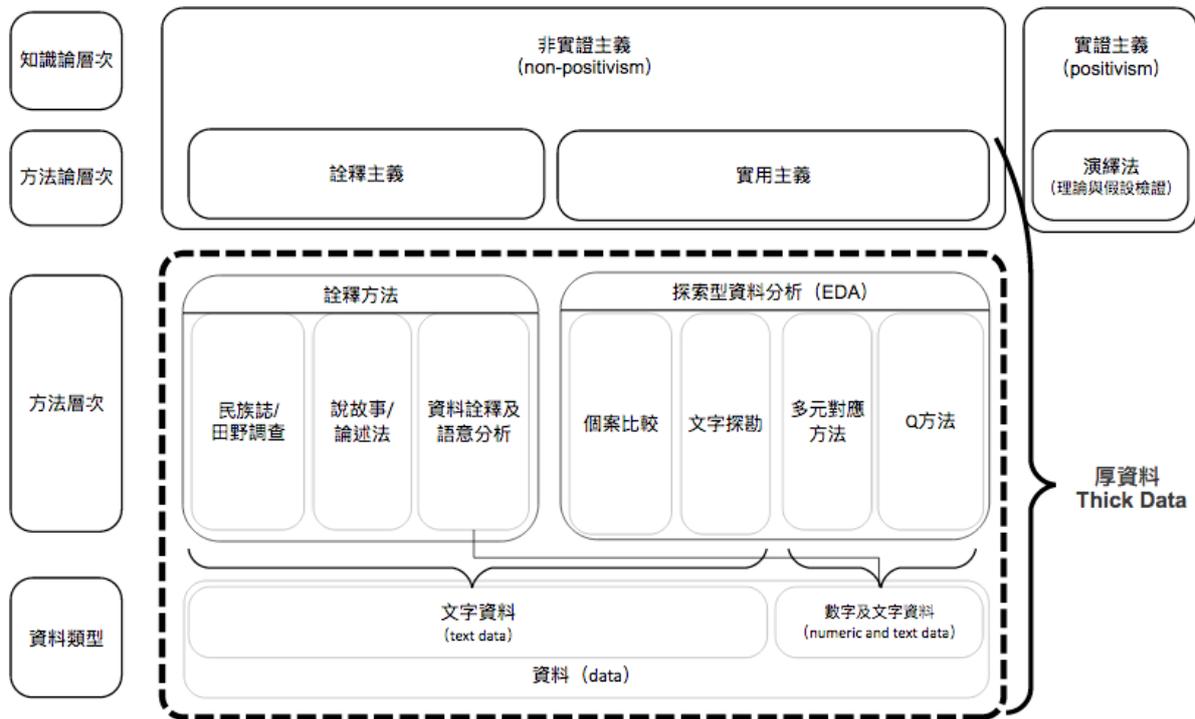
28)。由此來看，描述與詮釋都能產生意義，兩者在「產生意義」的目的下，可以分也可以合。無論是單就資料的描述來產生意義，還是在描述後加以系統性的詮釋來產生意義，都有其價值。不過，厚資料這個途徑與詮釋學派不同之處，在於「資料本身是客觀存在」的觀點：厚資料途徑下，針對客觀資料的詮釋，並非完全主觀，而是相對主觀。若是資料得以客觀地存在，那麼即使不同研究者各自的主觀的詮釋也有其邊界，甚至在資料面前，相對的主觀詮釋也能產生客觀的知識。

我把這個「探索+描述+詮釋」的求知立場，當作是一個知識建構的方法論，將這個隱藏在社科不同學門的研究方法之中，得以串連多個不同研究方法的方法論立場叫作「厚資料途徑」，也就是：

探索 + 描述 + 詮釋 = 厚資料方法論 (a thick data methodology) 或厚資料途徑 (a thick data approach)。

這個方法論中，質性與量化的區隔必要性不高。例如調查法與焦點團體法在傳統上分別被稱為「量化」與「質性」方法，但是在這個厚資料途徑的視野下，都可以為意義發掘這個目的所用。這種具詮釋學研究的厚資料觀點，便可以做出與以理論導向的實證研究傳統不同的貢獻。以驗證理論為目的研究，基於理論承載的要求，資料用於理論檢證或是粹取理論，而非意義的開發、描述及詮釋。以民意調查為例，由於不斷受到樣本代表性困擾（例如手機普及、行為意義的忽略、民調資料私有化、資料視覺化技術未成熟），故除非持續擴大採樣對象以及加權外使其較有代表性，否則難以有效驗證理論 (Dillman, Smyth, & Christian, 2014; Lee & Valliant, 2009; Mutz, 2011; Paolacci & Chandler, 2014; Paolacci, Chandler, & Ipeirotis, 2010; Romaniuk, Patton, & Carlin, 2014; Steinmetz & Tijdens, 2009)。但從厚資料觀點來看，若不強調代表性之後，資料也可能有其貢獻。實證主義的心理學透過小樣本結合實驗方法，仍推進了整個學門的進展；同樣的，在厚資料途徑下，即使是用於假設檢證的調查資料也可以出現以探索為目的應用。

圖一：厚資料途徑的知識地圖



知識論是方法論的根，而方法論的選擇會影響研究方法的選擇 (Hay, 2002)。以下，我進一步從方法論層次與研究方法層次，進一步勾勒厚資料途徑這個系統。由「厚資料途徑的知識地圖」(圖一)可以看出，厚資料作為一個方法論，包括了知識論層次的非實證主義、方法論層次的詮釋主義與實用主義，以及方法層次的詮釋與探索型資料分析方法。這些方法所對應的資料類型則包括了數字及文字資料。

知識論層次來看，實證主義是當前的學術主流，以理論建構與驗證（亦即「以原理解釋現象」）為求知的目標。在非實證主義的知識論立場中，實用主義「不帶著理論目的進行探索」、「解決問題的行動即是知識」的求知立場，與詮釋主義的「論述和意義即知識」的立場，與實證主義扞格不入，自然不受實證主義者所青睞。由於詮釋主義不強調探索、實用主義不強調描述與詮釋，因此厚資料方法論不獨依其中一支，而是兼容這兩個知識論立場。

就方法論及研究方法層次來看，目前並沒有一個名詞能同時含括以「探索+描述+詮釋」作為目的的方法論。我們所見，語言學和人類學的詮釋方法 (interpretation) 與統計學中的探索型資料分析 (exploratory data analysis) 最接近這個「厚資料方法論」的內涵。這兩種方法論各自包含一系列尚未「整隊」的研究方法：「資料詮釋」(data interpretation)、「說故事」(story-telling)/論述分析 (discourse analysis)，以及民族誌 (Ethnography)/田野調查 (field survey) 可統稱為詮釋方法。²「Q方

² 「論述分析」(discourse analysis) 是種捕捉語意的方法。近年以網路海量文本資料作為的論述分析資料來源的，則稱為數位論述分析(digital discourse analysis)。論述分析的方法，目的是主觀詮釋文本，以助讀者明白言說者的思路 (Gee, 2014a, 2014b; Page, 2014; Tannen & Trester, 2013; Thurlow & Mroczek,

法」(Q-method)、「多元對應分析」(Multiple Correspondent Analysis, MCA)、「文字探勘」(text-mining)、「個案比較法」(case comparison)等均是屬於樣態識別(pattern recognition)的方法,可統稱為「探索型資料分析方法」。

至於這些研究方法所對應資料類型,則如上所說,包含了文字資料(textual data)以及數字資料(numeric data)。這些研究方法均可處理文字資料,而以民意調查資料為分析對象的多元對應分析法及以訪談資料為分析對象的Q方法,則可同時文字及數字資料。至於其他類型的資料(如照片、錄音、影像、動漫、音樂等型態的資料)仍然可視為是可運用「厚資料途徑」發掘意義的礦脈,包括的範圍從一般民意調查、企業財務報表、新聞報導、政府公報、政治開放數據(open data)到政治人物言談、群眾意見、行為數據等,已含概絕大多數的資料型態。若我們把文字或所有的數位化資料比喻作礦脈,那麼厚資料途徑下的各種研究方法便可視為是開採這個意義礦脈的工具。在這個方法論體系中,研究者不只滿足於所探勘出的礦石(資料),而同時在乎礦石中的貴金屬(意義)。例如使用Q方法與多元對應分析法的研究者會從受訪者意見中匯整出意見的樣貌與類型;使用文字探勘的研究者會呈現透過詞頻分析的統計技術找出特定或不特定詞彙出現頻率以及關鍵詞連結叢集,辨識出特定概念發展與變遷;使用個案比較法的研究者著眼於如何透過個案比較,捕足現象(如民主化)的發展模式;使用詮釋研究方法的研究者將以「說故事」或田野觀察論述分析如何勾資料或現象背後的意義。

三、厚資料觀點下的民調資料收集

2011-2013年我擔任中山大學民意與市場調查中心主任期間,看到許多當前以電話調查進行研究的限制,並開始構思及籌畫一個網路調查平台,作為與電訪業務的互補。

當前電話調查訪問的限制包括了(1)電訪成本節節上升到一般研究者負擔不起;(2)一般電訪單位所訓練出的學生因為忙於電訪操作的細節而對資料分析的市場脈動失去靈敏嗅覺;(3)電話調查的觸及或含蓋率節節下降,連它能宣稱的樣本代表性也已失色;(4)因為「產學合作」,電訪資料清一色被委託單位私有化,無法釋出供民眾取值取用;(5)往往因委託者的需求、成本以及電訪問卷長度的限制,問卷的內容偏向淺碟化,既無法深入問到問題的核心,又無法讓受訪者從訪問中得到什麼收穫。我當時並無足夠的資源、時間,來面對解決這些令我擔憂的困境。

令我擔心的,除了電訪業務及學生訓練無助於學生累積就業能力之外,還有以下幾點:(1)新聞報導常出現錯用或濫用民調知識「製造」錯誤或引導式的新聞;(2)網路調查根本沒有母體也無從抽樣,但使用者和報導者卻常魚目混珠,讓專業沉淪;(3)民眾對民調普遍反感,因為調查項目往往是以委託者為出發點,為委託者服務,所以政策、選舉和商業活動之外幾乎其他型態的調查內容存在;(4)民調產業面對正在起飛中的大數據應用,以及開放資料(open data)思潮無動於衷,甚至到了2013年,即使是政府大力推動開放資料,也還未到把「取之於民該還於民」的政府民調資料,放入應該同步推動開放的清單之中。更不用說立法院尚未正視這個「民調資料透

2011; Wortham & Reyes, 2015)。

明化」的理念，推動立法，讓公民取得資料問卷了。我在 2011 年已略嗅出這些問題，也覺得這是個需要先有磚才有玉的事業，所以我就先自己來燒磚。

在教育部及科技部計畫的經費支援下，我籌組了師生團隊，建置了「微笑小熊調查小棧」(smilepoll.tw) 平台，於 2012 年上線，在非常有限的資源協助下，運作至今。隨著參與人數逐漸增加（註冊會員超過 6,000 人，每超過 20,000 人次填寫問卷，累積出的且釋出的網路調查資料已超過 120 筆）。目前每月提供 3 到 5 筆不同主題的問卷供民眾參與，每份調查參與人次約為 600 人。小棧透過抽獎活動控制成本，但也已成功吸引許多非會員民眾加入，並及成功驅動會員持續參與網站的問卷調查的興趣。我將收集到的問卷調查報表與資料檔在兩週內發佈到網站上供會員登入官網後閱覽與下載（如圖二所示），並將分析結果同步發布於臉書粉絲專頁及部落格。頁面上會記載每份調查名稱、時間、參與人數，以及各資料下載的情況。民眾不需經過繁瑣的索取的程序即能自由下載。

圖二、smilepoll.tw 的開放資料

問卷編號	問卷名稱	開始日期	結束日期	參與人數	調查結果	下載次數	原始資料(.csv)	下載次數	問卷題目	下載次數
1093	呼吸多健康?	2015-11-27	2015-12-11	720		7		3	-	-
1092	我的陸客印象	2015-12-19	2016-01-15	983		10		1	-	-
1091	消費實折扣	2015-11-20	2015-12-04	701		7		1	-	-
1090	運動由你定義	2015-11-06	2015-11-20	708		14		4	-	-
1088	兩岸關係之我見	2015-11-20	2015-12-14	824		15		1	-	-
1087	小熊看電競	2015-10-30	2015-11-13	948		12		1	-	-
1086	小熊看教育界的數位落差	2015-10-23	2015-11-06	727		6		1	-	-
1085	小熊看金馬獎	2015-10-10	2015-10-23	711		11		4	-	-
1084	小熊看國家認同	2015-10-18	2015-11-06	911		24		4	-	-
1083	小熊看長青世代	2015-10-02	2015-10-16	703		14		4	-	-

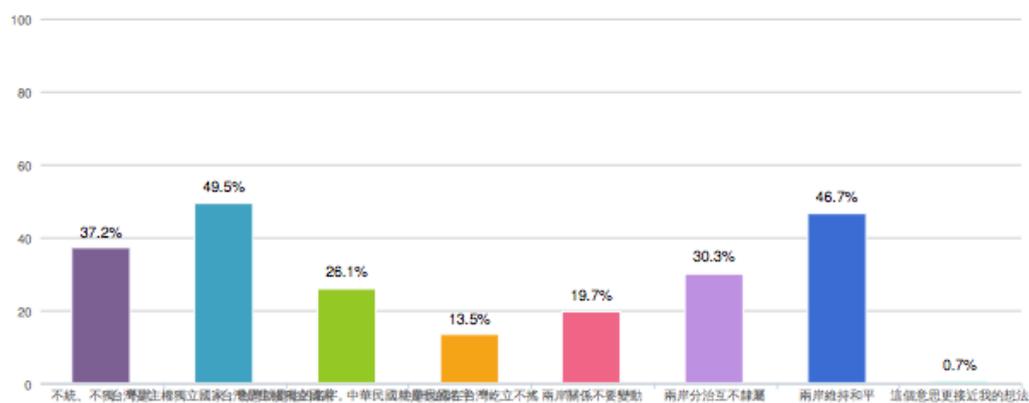
小棧作為厚資料途徑理念的實踐，除了提供學術上的方便抽樣（convenient sampling）的資料外，其受訪者可被反覆接觸的特質才是最珍貴、最有價值的地方。這個社群的存在及持續擴張，將讓學者得以打破研究方法量化質化的界限，透過網路問卷直接進行質性訪談，或從線上走到線下進行約訪，更可以透過串連資料庫的方式研究態度或價值變遷，以及透過實驗設計進行理論的驗證。換言之，「微笑小熊調查小棧」象徵的，是能讓一群不完全為物質利益而參與問卷、隨時反映想法，支援學術調查的自由理想成真。「微笑小熊調查小棧」的存在，也開啟了社會科學學門、人文學科、管理以及資訊管理學門真正產生合作的可能。它的重要性不再只是「跨」學門（cross-

discipline) 而是以問題導向 (problem-driven) 所產生出的學門消解融合 (synergy of disciplines)。

這個平台所能進行的調查型態包括了一般意見調查及 A/B 分群實驗，但最特別的是它能隨著研究者的需要，進行質性的調查。舉例來說，一般的電訪問卷只能問到受訪者的統獨偏好，但 smilepoll.tw 允許研究者進一步探詢每個偏好的意義。如圖三所示，受訪者得以在沒有壓力的狀況下，自由勾選或書寫對於維持現狀的想法，呈現出豐厚多元的偏好。研究者甚至能進一步 (在知情同意前提下) 約訪這些會員，讓線上和線下的訪問及意見探索的研究不再有隔閡。

圖三、質量合一的民意調查舉例 (維持現狀)

39. 下列那個維持現狀的意思最接近您的想法? (可複選)



Value	Percent	Count	Statistics
不統、不獨、不武	37.2%	366	Sum 9,258.0
台灣是主權獨立國家, 台灣就是他的名字	49.5%	487	Average 9.4
台灣是主權獨立國家, 中華民國就是他的名字	26.1%	257	StdDev 5.4
中華民國在台灣屹立不搖	13.5%	133	Max 92.0
兩岸關係不要變動	19.7%	194	
兩岸分治互不隸屬	30.3%	298	
兩岸維持和平	46.7%	459	
這個意思更接近我的想法	0.7%	7	
Total		983	

Responses "這個意思更接近我的想法"	Count
Left Blank	976
中國是台灣的兄弟之邦	1
中華民國是主權獨立國家, 新台幣印有"中華民國"能在世界流通! 讚!	1
中華民國治理臺灣金馬	1
兩個中國	1
可以統一但要自由民主才行	1
我們要走向進步, 獨立, 民煮的臺灣	1
相依相扶	1

四、代結論

目前國際知識社群才剛始在「厚資料方法論」進行論述，因此這個論述具有集結國內學者、面對國際，甚至在新興議題上取得國際上發言權的潛力。就知識建構來看，非實證主義的許多研究方法目前仍散在不同學門中，需要更有系統的互動與激盪，來產生如實證主義一樣的知識生產力道。具體來說，我期待這個論述有助於將長年持有相同知識論立場（例如都採歸納途徑或是詮釋途徑進行研究），但不一定相識的學者，開始在方法論層次進行交流。雖然本文是從非實證主義的知識論出發，但並不是為了排斥實證主義而寫，而是站在知識論的多元主義立場（epistemological pluralism）期待看到更多實證主義與非實證主義之間的對話，進而減少非必要的批判與相輕。

由知識論的多元主義立場帶出的是「方法論多元主義」（methodological pluralism）。採取厚資料途徑的最大優點在於，從這個視野所看出去的求知世界，諸多為了理解而存在的研究方法之間沒有隔閡。一些既存但已長年未受到重視的研究方法將得以有一個新的姿態重新站上研究的舞台。例如「Q方法」、「多元對應法」、「個案比較法」等。至於「資料詮釋」和「說故事」等由下而上的資料及其意義探索方法，也將變得值得政治學者重視與學習。社科學者的研究工具箱在厚資料途徑視野下將變得更大更豐富。

從資料性質來看，厚資料途徑視「資料」（data）為「意義的礦脈」（mine of meanings），讓資料由客為主，提供了實證主義「資料是手段，理論才是目的」之外的選擇。從「探索+描述+詮釋」的方法論來看，用於探索、描述與詮釋的「資料」的型態無限寬廣，包含了質性（像是民族誌或圖片影音）與量化的資料（例如文字資料庫和民調資料），因此傳統的質化量化研究方法的二分法，甚至是方法論上的對立，將會被淡化，變得沒有必要。也就是說，在這個厚資料的觀點下，資料使用者不必硬分質性還是量化，未來的研究者及決策者將會把重點移到在能夠從中資料中挖掘出多少豐富的意義，為此引介或開發新的資料分析方法。

第三，從人才培育來看，厚資料途徑下所看到的政治學，是一個全新的看待人類社會生活資料的觀點。厚資料途徑在意義產出過程中包括了收集、分析及解讀三個環節，因此厚資料觀點下的政治學研究訓練將強調：

1. 資料的創造力：創造或收集具有豐富意義的資料的能力（如設計具有深度的問卷）。
2. 資料的分析力：發展及學習挖掘意義的方法，也就是經由訓練來自行分析並且正確解析出意義礦脈的能力。
3. 資料的詮釋力：看出被挖掘出來的意義並正確應用到治理、洞悉問題上及預見潛在問題的能力。

即使社科界追不上（也不必追）產學界對「資料科學」及「資料科學家」的概念及產業形塑工程，政治學界仍有足夠的潛力開創新的路徑。資料科學已短時間內成為結合大數據、資訊工程、統計學、圖像視覺化等領域的新興的新產業。不過，未來的資料科學將不只需要資料科學家（data scientists）、資料新聞專家（data journalists），也將愈來愈需要資料鑑賞家（data story-tellers）。若能在此戰略構想上進行新一代社科

人才的培育，政治學將有機會重新得到社會各界的重視。「大數據」這個概念已經將資工、資管、財金、應用數學、統計人才整合起來；「數位人文」已將歷史、哲學、文學、社科與資訊科學整合起來；現在，「厚資料」將足以將從事社會、政治、經濟、社會、教育、心理、企業管理、財務金融、傳播行銷、新聞視覺化等領域的學者融合在一起，產官學界將可以依此形塑全新的合作概念。

參考文獻

- 項潔、涂豐恩。2011a. 〈導論—什麼是數位人文〉。載於項潔主編。《從保存到創造：開啟數位人文研究》：9-28。臺北市：國立臺灣大學出版中心。
- 江佩純, & 項潔。2014。《數位人文研究與技藝》。臺北市：國立臺灣大學出版中心。
- 項潔。2011a。《從保存到創造：開啟數位人文研究》。臺北市：臺灣大學出版中心。
- 項潔。2011b。《數位人文在歷史學研究的應用》。臺北市：國立臺灣大學出版中心。
- 項潔。2011c。《數位人文研究的新視野：基礎與想像》。臺北市：國立臺灣大學出版中心。
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method* (4 edition). Hoboken: Wiley.
- Gardiner, E., & Musto, R. G. (2015). *The Digital Humanities: A Primer for Students and Scholars*. Cambridge University Press.
- Gee, J. P. (2014a). *An Introduction to Discourse Analysis: Theory and Method* (Fourth edition). New York: Routledge.
- Gee, J. P. (2014b). *How to Do Discourse Analysis: A Toolkit* (Second Edition). Milton Park, Abingdon, Oxon: Routledge.
- Geertz, C. (1973). Thick description: Toward an interpretive theory of culture. In *The Interpretation of Cultures: Selected Essays* (First Edition edition, pp. 3–30). New York: Basic Books.
- Gold, M. (2012). *Debates in the Digital Humanities*. (M. K. Gold, Ed.).
- Gray, E., Jennings, W., Farrall, S., & Hay, C. (2015). Small big data: Using multiple data-sets to explore unfolding social and economic change. *Big Data & Society*, 2(1), 2053951715589418. <https://doi.org/10.1177/2053951715589418>
- King, G., Pan, J., & Roberts, M. (2013). How censorship in China allows government criticism but silences collective expression. *American Political Science Review*, 107(2 (May)), 1–18.

- Lee, S., & Valliant, R. (2009). Estimation for volunteer panel web surveys using propensity score adjustment and calibration adjustment. *Sociological Methods & Research*, 37(3), 319–343. <https://doi.org/10.1177/0049124108329643>
- Madsbjerg, C., & Rasmussen, M. B. (2014, March 21). The power of “thick” data. *Wall Street Journal*. Retrieved from <http://www.wsj.com/articles/SB10001424052702304256404579449254114659882>
- Marsh, D., & Stoker, G. (2002). *Theory & Methods in Political Science* (2nd ed.). Palgrav Macmilan.
- Mutz, D. C. (2011). *Population-Based Survey Experiments*. Princeton University Press.
- Page, R. E. (2014). *Researching Language and Social Media: A Student Guide*. London ; New York: Routledge.
- Paolacci, G., & Chandler, J. (2014). Inside the turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, 23(3), 184–188. <https://doi.org/10.1177/0963721414531598>
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on TAmazon Mechanical urk. *Judgment and Decision Making*, 5(5), 411–419.
- Rasmussen, M. B., & Hansen, A. W. (2015, November 16). Big data is only half the data marketers need. Retrieved February 13, 2016, from <https://hbr.org/2015/11/big-data-is-only-half-the-data-marketers-need>
- Romaniuk, H., Patton, G. C., & Carlin, J. B. (2014). Multiple imputation in a longitudinal cohort study: A case study of sensitivity to imputation methods. *American Journal of Epidemiology*, 180(9), 920–932. <https://doi.org/10.1093/aje/kwu224>
- Schreibman, S., Siemens, R. G., & Unsworth, J. (Eds.). (2016). *A New Companion to Digital Humanities*. Chichester, West Sussex, UK: John Wiley & Sons Inc.

- Steinmetz, S., & Tijdens, K. (2009). Can weighting improve the representativeness of volunteer online panels? Insights from the German Wage indicator data. *Concepts and Methods (Newsletter of the IPSA Committee on Concepts and Methods)*, 5(1), 7–11.
- Tannen, D., & Trester, A. M. (Eds.). (2013). *Discourse 2.0: Language and New Media*. Washington, DC: Georgetown University Press.
- Tencer, K. (2014, May 9). Why big data only tells half the story. Retrieved February 13, 2016, from <http://www.theglobeandmail.com/report-on-business/small-business/sb-managing/why-big-data-only-tells-half-the-story/article18541573/>
- Thurlow, C., & Mroczek, K. R. (Eds.). (2011). *Digital Discourse: Language in the New Media*. Oxford ; New York: Oxford University Press.
- Uprichard, E. (2015, February 12). Most big data is social data: The analytics need serious interrogation. Retrieved from <http://blogs.lse.ac.uk/impactofsocialsciences/2015/02/12/philosophy-of-data-science-emma-uprichard/>
- Wortham, S. E. F., & Reyes, A. (2015). *Discourse Analysis Beyond the Speech Event*. London ; New York: Routledge.