

大數據分析作為一種社會現象：批判性的觀點

李俊毅

一、引言

本文嘗試探討大數據(big data)及相關科學的興起、許諾、以及背後的政治性。此一做法的目的係著眼於當前我國在論及大數據時，往往僅將它視為價值中立的技術或科學，也因此對於銳意發展大數據及相關科學的政府與企業而言，他們往往強調大數據可帶來的創新與利益。舉例來說，在科技部 106 年度的「數位經濟前瞻技術研發與應用專案計畫」中，即主張「大數據既龐大且複雜，其內之『知識』擁有有非常大的價值」；此一計畫的目的，從而是「探討大數據之關鍵智慧分析與處理技術，以及其可行高附加價值之創新應用，以有效提高國內相關產官學界於大數據分析帶來之價值」(科技部，2017)。相應的，持保留態度者，則多強調大數據分析及應用對於個人隱私權的可能侵害。本文並不擬對大數據的正反意見做出評判，而旨在指出，大數據(及其他新興科技)的興起宜置於當代的經濟與社會脈絡觀察，而由此來看，正反的觀點同時反映當前新自由主義的特色，因此要為大數據之利弊作出裁決的嘗試，終將是徒勞的。更值得關注的議題，毋寧是「數據」或「資料」的意涵、特質、以及當這些數據能更大量與快速地被生產、收集、處理與分析後，對既有社會生活的影響。

二、自由主義下的科技與社會

一般咸信，科技的發展有助於增進國家與社會的福祉；以此，當代國家無不關注科技的發展。事實上，國家與科技的關係是更密切的：科技的進展不僅仰賴國家資源的挹注，近代國家的形成更是有賴諸多科學與技術的發展與應用。法國歷史哲學家傅科(2007, 2008)追溯政府一詞的意義，認為直至 18 世紀之前，治理(to govern; *gouverner*)一詞被廣泛的使用在自我管理、引導家庭與小孩、操持家務、甚至引導靈魂等面向，而不專指國家或行政。由此，傅柯將「政府」一詞定義為「行為的引導」(the conduct of conduct)，也就是關於人口的行為之引導，包括對自己的治理以及對他人的管理，而不專指狹義的政治結構或國家的經營。「政府」因此包含多種管理當局與機構，而不是一個權力的中心。透過這些場域、技術與知識，政府藉由運作人們的慾望、抱負、利益與信念等，而達到改變或引領人們行為的目的。重要的是，此種權力的運作，對象不僅是他人，也包含自我的管理。

把治理性的觀點歷史化，傅柯認為西歐在 16 到 18 世紀這段期間，對於治理的看法出現轉折：治理不再是依循諸如自然法或宗教，也不再依靠統治者個人的權術與智慧，而需要援引各種知識，以掌握關於國家的具體事實，以便對內促進發展，對外在國際競爭中保護國家。18 世紀起，「人口」(population)一眾多的個人，其和所處的環境構成生物性的連結一開始出現。人口成為一種現象或事實，有其規律、特質、內在邏輯等，是國家可以知道且介入的。由於人口的概念伴著「流通」一活動、交換與接觸，政府的功能便是確保流動能順利運作，以極大化流動(如商業)帶來的利益，並抵消其負面的影響。由此而來的，則是「偶然性」(contingency)的概念，因為流動使事物與事件的發展變得無法完全知道與控制；面對未來的不確定性，「安全」亦變得重要。簡言之，安全即是確保流動的利益極大化與風險極小化。由此，安全與自由是相構成的：自由，無論是人或市場的流動總是伴隨著風險與不確定性，安全的機制於是必要；隨著自由的程度越高，對安全的需求就越高，安全機制的發展也就越繁複。自由與安全，或「更少的治理」與「更多的治理」(Neumann & Sending, 2010: 3)，因此成為當代政治生活兩大構成原則的邏輯。

傅柯認為 18 世紀起的治理性，可以用自由主義(liberalism)稱之。在此自由主義並不是指一套學說或意識形態，而是政府為促進人口的福利、壽命、健康與財富時的施政的方法或工具。自由主義最主要的特色是：國家政府藉由透過某些自由的活動，以及培養被治理的人們適當的自我管理習慣以盡可能的延伸其治理。18 世紀末開始，「社會」(society)的概念開

始出現，並如同經濟一樣，被認為是個具有特定規律與事實的場域，而政府若對之進行過多的治理，則它事實上將無法治理。社會科學的出現，便可看成是關於「社會」此一事實的技術，也就是試圖理解它本身的變遷，特別是政府/政治適不適合對之進行干預。而「市場」的概念在此扮演示範的角色，因為它最能彰顯過度治理的不良結果。市場是去中心的，處於市場中的個人，在一段時間後，將透過和他人的互動而改變其個人內在的行為標準，例如學會謹慎、勤奮、守時與自律等；表現在國家層次，則文明人口所處的國家，也被期待將以相似方式，和平地與有秩序地互動。換言之，市場有「文明化」的功能，講求法治(國內)與條約(國際)並促進商業，因此是當代國家治理內部人民與經營彼此關係的方式，也就是「透過自由來治理」(governing through freedom)。

科學(包含人文與社會科學)與技術的發展，在此扮演重要的角色。民族主義學者 Benedict Anderson (1991: 6)指出，「任何大於由[成員間]面對面接觸所組成的原初村落之社群都是想像的」，凸顯出人在認識能力上的限制。由此延伸，近代國家要知道關於其人口或所轄社會之「事實」，絕非簡單的工作；它除了需要國家發展專門的官僚機構，也需要各種學科的分殊與專業化。在眾多科學中，「統計」(statistics)的發明尤為重要，其字根甚至意味「國家的科學」(Foucault, 2007: 354)。統計學提供了一套方法或技術，使人們能從相對少數的樣本，推估整個母數或全體的情況；對國家而言，統計因此使國家知道其統治客體的各項事實。到了當代，統計的應用，則從(看似)單純地呈現事實，延伸至引導乃至管理公眾的行為(如民意調查的技術)(cf. Sætnan, Lomell & Hammer, 2011)。在統計學之外，機率、數據科學與演算法等，也都是「取得」關於真實世界的資料之技術。由此而來的結果則是，當國家對人口及個人的事實掌握越多，它便越能介入或干涉人們的社會生活。當前國家與個人的關係早已不僅限於納稅、服兵役、受義務教育等權利與義務關係；國家更透過專家與知識，告訴人們什麼是安全、健康、幸福、與成就，而我們也因為國家龐大的職能，而日益要求各式安全的保障與福祉的提供。

質言之，科技的發展，使國家有能力知道並掌握其人口或社會的事實，從而能一方面鼓勵並享有人員、貨品、資訊與資本因自由流動而帶來的好處，另一方面確保自由流動帶來的不確定性與危害能受到控制。相較於之前的科技發展(如統計)，大數據分析固然在性質上有所不同，但依循前述邏輯，我們已可看到何以關於大數據的論辯不會終止：它首先許諾我們對於社會的整體與個體。都能知道更全面且更細微；其次，在此基礎之上，我們可發現與發展各式有益於個人、企業、乃至於社會福祉之措施，也可藉此防範各種不安全與危害；最後，對「更多的治理」之需求，也相應地帶來「更少的治理」之聲浪，國家的知道(knowing)能力是否應有界線、此一界線該如何劃分、由誰劃分與掌控，也是由此而生的必然問題。

三、大數據的社會性與研究趨勢

由以上的立論來看，對於大數據及相關分析技術的正反論辯，是一個持續發生的社會現象。深化或推進此一辯論的方式，是探討數據或資料的社會乃至政治性。資料或數據並非對事物的性質之簡單陳述，而是透過特定的方式生產、搜集與處理的結果。以統計來說，統計常被視為對社會的簡單、直接與客觀之描述，但人們選擇計算(count)什麼、不計算什麼、由誰來計算、賦予這些計算何種類別與價值，卻可以造成重大的差別。數字的生產及其意義的賦予，因此帶有深刻的社會與倫理啟示，並往往是政治性的(Fioramonti, 2014; Mair, Greiffenhagen & Sharrock, 2016; Sætnan, Lomell & Hammer, 2011)。與此呼應，王甫昌(2005)對戰後台灣戶口普查籍別類屬的研究亦指出，雖然概念上普查的對象都是台灣的家戶或人口，但在此過程中，由「省籍」到「族群」作為劃分身份的依據，反映了從「代表中國的政治需要」到「以台灣為範圍的主體想像」之政治與社會變遷。這類的研究顯示，知識並不是「發

現」真理，而毋寧是生產某些真理，從中特定的主體與客體能被建構，並成為權力介入的空間與對象(Krasmann, 2017: 11)。

將此視角應用於大數據分析，亦即從批判性的立場探討大數據的現象與可能啟示，迄今仍是個相對新興且具挑戰性的研究領域，因為數據的生產涉及編碼、演算法、物聯網、人機關係等軟、硬體科技議題，需要跨領域的對話與合作(Aradau & Blanke, 2015)。既有文獻在研究議題與方法上迄今仍是多元與發散的，本文以下將由大數據的定義帶出幾項討論。

儘管大數據已是人們琅琅上口的用語，當前對「大數據」的定義仍莫衷一是。比較主流的定義，是環繞在「3V」的概念上增添。「3V」意指容量(Volume，即大量資料的使用)、多元性(Variety，資料來源的分散，資料可儲存在分散的結構或甚至是非結構化的)以及速度(Velocity，資料分析或處理的速度往往是即時的)。在此基礎上，另有論者將可信度(Veracity)、價值(Value)、虛擬(Virtual)等當成大數據的特色(Broeders, Schrijvers & Ballin, 2017: 5-6)。前述科技部「數位經濟前瞻技術研發與應用專案計畫」，即是以容量/體積、速度、多變性、可信度以及價值等五個特色界定大數據。

這些界定方式固然有助於理解大數據的特質，卻無益於解釋其之普遍化乃至於「誤用」。大數據分析的流行，應被看成是一個文化、技術與學術現象，是技術、分析與迷思/神話三者的互動(boyd & Crawford, 2012: 663)：

- (一) 技術：極大化運算能力以及演算的精確性以收集、分析、連結與比較大型的數據集(data sets)。
- (二) 分析：在大型數據集的基礎上識別特定的模式，藉此做出經濟、社會、技術與法律的宣稱。
- (三) 迷思：認為大型的資料集能提供更高形式的情報與知識，而能生產前此無法得出之洞見，這些洞見更被賦予真理、客觀與精確之光環。

本文認為，當前對大數據的批判性探討，或可藉由這三個層面予以組織。其中，由於「迷思」的訴求對既有之科學與社會生活衝擊較大，引起最多的關注。此一迷思的威力，在於相較於統計學須以樣本為基礎，大數據以其資料量的龐大，讓人們產生大數據之樣本數即等於全體(N=all)的信念(Hildebrandt, 2013)。此一“N=all”的宣稱之所以可能，則是立基於個體資訊的數據化。因此，大數據分析許諾的是我們不僅能藉此知道社會整體，也能藉由個體在真實與虛擬世界中因互動而留下的種種蹤跡發現其行為模式。循此，科學似乎不再需要追求因果關係，而可以在憑藉大數據分析找出相關性。對社會生活來說，在「全球化」的趨勢下，人員、資訊、資本、與貨物的自由流動也帶來諸如恐怖攻擊、組織犯罪、毒品、走私、人口販賣、傳染病、乃至(非法)移民等問題，使國內與國際秩序似乎隨時處於「不安全」或「危機」的狀態。在“N=all”的前提下，輔以數位工具的發展，大數據分析在「稻草堆中尋找一根針」就不再是遙不可及之事；而當大數據分析越能宣稱在安全保障上的功能，就越能合理化情報與安全機構的擴充，以及其更大規模的搜集與處理個人資訊的做法(Aradau, 2015)。若大數據分析在「除弊」方面—防制人與事物因自由流動而引起的安全問題—具有吸引力，那麼其在「興利」方面—疾病的治療、災難的預警與防治、對消費者之喜好新的掌握、客製化服務的提供等—的前景，就更易令一般大眾接受。

在相當程度上，大數據的迷思來自於其提供預測未來的許諾。此一許諾之所以動人，反映了人們對真理或事實的追求；當我們能知道整體事實，就能進一步對人與事物施加作為以改變、調整、或改善之。「追溯我們物種的行為，『真理的表達』(true utterance)是一個幾乎難以匹敵的力量...從很早開始，真理對人們而言就是最具效力的言語武器之一」(Dumézil, 1943: 243-244, cited from Foucault, 2013: 84)。

在「迷思」的層面上，持批判性的文獻大致有兩項回應方式。其一是從政治或哲學的角度，探討大數據分析對社會生活的影響。例如 Hildebrandt(2013)認為我們不至於落入被大數據分析或機器控制的局面，因為大數據分析僅能數據化可以被化約為數據的生活層面，因此“N=all”事實上不可能存在。然而她亦提醒，人之擁有自由，乃因人類是複雜的動物，能不斷賦予自己乃至世界新的意涵，因此人們的思考與行為方式是否會因在數據化的普及而趨同或簡化，方是關鍵所在。Ruppert, Isin & Bigo (2017)則嘗試為數據、知識與權力的關係提出研究綱領。數據涉及了事物(伺服器、設備與纜線等基礎設施)、語言(編碼、程式與演算法)、以及人(科學家、企業家、工程師、資訊技術人員、設計師)等，並建構了新的世界，在其中出現了新的主體(行為者)以及權利的宣稱。因此唯有理解這些可能性條件的運作，人們才有介入並形塑資訊政治的空間。當然，此一取向的研究並不是大數據分析出現之後才有，亦不僅限於大數據分析，而可視為是更廣義的科技與社會(或政治)之研究的延伸(cf. Amore & Piotukh, 2016; Berry & Fagerjord, 2017; Desrosieres, 1998; Fioramonti, 2014)。

其次，有鑒於大數據分析在國際與國內政治的流行多與「安全」有關，許多文獻乃關注大數據是否及如何改變國家機器和公民之間的關係。2001年九一一恐怖攻擊事件以降，美國與部分歐洲國家以反恐為名，加強了對國際移動人口(移民、難民、與一般旅客等)之監視，也連帶合理化其對本國人民之資料的搜集、分析與處理。若干學者因此探討「風險」概念對治理的影響(Amore & de Goede, 2005; Amore, 2013; Van Munster, 2009)、安全治理的技術如生物辨識系統與各式資料庫的建置(Vaughan-Williams, 2015)、以及由此而生的監視問題(Aradau, 2015; Aradau & Blanke, 2015; Monahan, 2016)。大致來說，這些研究並不探討大數據分析及相關技術的「成效」，因為概念上，若風險管理與監視技術有效運作，則恐怖攻擊的事件自然大幅度減少甚至不會發生，但吊詭的是，零星的事件難以證明安全措施之足夠，反而常常是相關機構強化其監控與資料取得的證明。它們因此關注日益「安全化」的世界之形成，以及此一變遷對國家與公民之間應有的距離之侵害。

在「迷思」的層面之外，近期亦有學者開始關注「分析」與「技術」的問題。Kitchin (2017)指出演算法是一組明確界定的步驟，以結構化的方式處理數據而生產特定結果，其本質是偶然的(contingent)、個體發生的(ontogenetic)與表演的(performative)，因此應該在社會與科技的脈絡中探討其實踐。Aradau & Blanke (2015)則從美國關於大數據的法律爭議，主張我們對數據的理解，應該持續與資訊科學的發展對話。例如雖然許多人已可接受諸如「沒有原始資料(raw data)的存在」、「原始資料亦是資料」等看法，但美國的相關辯論已轉向「資料」與「後設資料」(metadata)是否有別。除此之外，支持大數據分析與人工智慧(artificial intelligence, AI)者常主張相關之分析大部分是由機器自動化處理，其中僅有很小的部分牽涉人工，因此大數據分析不致影響人權與隱私。但從 AI 的發展歷程來看，演算法的建構在許多面向上仍仰賴人為的模型化，因此大數據與 AI 涉及之人機關係，其界線非如安全專家所述之明確與遙遠。最後，資料與演算法之間的關係亦值得關注，蓋雖然大數據分析的倡議者主張只要資料夠多，就越能實現“N=all”以及在「稻草堆中尋找一根針」，但在演算法仍相當倚賴機率的前提下，數據越多是否能得到更多與更明確的正相關，或是產生更多錯誤，則值得進一步探討。上述研究取向將注意力從大數據分析的社會與政治啟示轉向數據與演算法本身，但亦因數據科學的專業性、數據與演算法的不透明性、以及數據及演算法本身的多元性，而訴求跨領域的對話與合作。

四、結語

本文嘗試理解大數據之概念及其應用的普及，並探討大數據的社會與政治分析之發展方向。將大數據分析的興起界定為一個社會現象而不只是一個工具或技術，則此一概念的流行反映或捕捉了我們對「事實」與「真理」的追求，因為自理性時代以來的一個基本信念是，唯有理解與掌握世界的態樣與規律，方能對其施加作用，而將世界導向更理想的狀態。

此一對知識或知道的意志，伴隨著國家自近代以來的形成而與政治權力產生密切的連結；國家藉由知道其人口與社會的事實，從而提供安全、福利與福祉。大數據產生之“N=all”的迷思，一方面呼應了前述的知識與政治脈絡，另一方面也使其難以受到根本的否定或挑戰。

以此，對於大數據分析的社會與政治分析亦出現轉折。許多研究關注大數據分析及相關領域對於既有政治與社會秩序、人與國家關係、隱私權與自由等層面的影響，部分研究則開始探討演算法與數據的本質。後者嘗試將本來被視為客觀與技術領域的事物置於社會、政治、與文化的脈絡觀察，進而反思大數據的「迷思」。然而，有鑒於資訊科學本身的專業性，既有研究仍處於萌芽的階段，而訴求更多跨領域研究的投入。

參考資料

- 一、王甫昌，(2005)。由「中國省籍」到「台灣族群」：戶口普查籍別類屬轉變之分析。台灣社會學，第9期，59-117。
- 二、科技部，(2017)。科技部「數位經濟前瞻技術研發與應用專案計畫」106年度四大研究領域之描述，<https://www.most.gov.tw/most/attachments/49b22f2d-e4fa-4fcc-8066-87eb7e1450af>。
- 三、Amoore, L. (2013). *The Politics of Possibility: Risk and Security beyond Probability*. Durham: Duke University Press.
- 四、Amoore, L. & de Goede, M. (2005). Governance, Risk and Dataveillance in the War on Terror. *Crime, Law & Social Change* 43(2-3): 149-173.
- 五、Amoore, L. & Piotukh, V., eds. (2016) *Algorithmic Life: Calculative Devices in the Age of Big Data*. Abingdon: Routledge.
- 六、Anderson, B. (1991). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso, 2nd edition.
- 七、Aradau, C. (2015). The Signature of Security: Big Data, Anticipation, Surveillance. *Radical Philosophy* 191: 21-28.
- 八、Aradau, C. & Blanke, T. (2015). The (Big) Data-Security Assemblage: Knowledge and Critique. *Big Data & Society* 2(2): 1-12.
- 九、Berry, D. M. & Fagerjord, A. (2017). *Digital Humanities: Knowledge and Critique in A Digital Age*. Cambridge: Polity.
- 十、boyd, d. & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society* 15(5): 662-679.
- 十一、Broeders, B., Schrijvers, E. & Ballin, E. H. (2017). *Big Data and Security Policies: Serving Security, Protecting Freedom*. The Hague: WRR. Retrieved from https://english.wrr.nl/binaries/wrr-eng/documents/policy-briefs/2017/01/31/big-data-and-security-policies-serving-security-protecting-freedom/WRR_PB6_BigDataAndSecurityPolicies.pdf.
- 十二、Desrosieres, A. (1998). *The Politics of Large Numbers: A History of Statistical Reasoning*. Cambridge, Mass.: Harvard University Press.
- 十三、Fioramonti, L. (2014). *How Numbers Rule the World: The Use and Abuse of Statistics in Global Politics*. London: Zed Books.
- 十四、Foucault, M. (2007). *Security, Territory, Population: Lectures at the Collège de France, 1977-78*. New York: Palgrave Macmillan.
- 十五、Foucault, M. (2008). *The Birth of Biopolitics: Lectures at the Collège de France, 1978-79*. Basingstoke: Palgrave Macmillan.
- 十六、Foucault, M. (2013). *Lectures on the Will to Know: Lectures at the Collège de France 1970-1971*. Basingstoke: Palgrave Macmillan.
- 十七、Hildebrandt, M. (2013). Slaves to big data. Or are we? IDP. *Revista de Internet, Derecho y Política* (17). Retrieve from: <http://journals.uoc.edu/index.php/idp/article/viewFile/n17-hildebrandt/n17-hildebrandt-en>.
- 十八、Krasmann, S. (2017). Imagining Foucault: On the Digital Subject and “Visual Citizenship.” *Foucault Studies* 23: 10-26.

- 十九、 Mair, M., Greiffenhagen, C. & Sharrock, W. W. (2016). Statistical Practice: Putting Society on Display. *Theory, Culture & Society* 33(3): 51-77.
- 二十、 Monahan, T. (2016). Built to Lie: Investigating Technologies of Deception, Surveillance, and Control. *The Information Society* 32(4): 229-240.
- 二十一、 Neumann, I. B. & Sending, O. J. (2010). *Governing the Global Polity: Practice, Mentality, Rationality*. Ann Arbor: The University of Michigan Press.
- 二十二、 Ruppert, E., Isin, E. & Bigo, D. (2017). Data Politics. *Big Data & Society* 4(2): 1-7.
- 二十三、 Sætnan, A. R., Lomell, H. M. & Hammer, S., eds. (2011). *The Mutual Construction of Statistics and Society*. London: Routledge.
- 二十四、 Van Munster, R. (2009). *Securitizing Immigration: The Politics of Risk in the EU*. Basingstoke: Palgrave Macmillan.
- 二十五、 Vaughan-Williams, N. (2015). *Europe's Border Crisis: Biopolitical Security and Beyond*. Oxford: Oxford University Press.