

書評

面對巨量資料時代的海量思維： 《大數據》

陳啟清*

書 名：大數據 (BIG DATA)

作 者：Viktor Mayer-Schonberger and Kenneth Cukier；林俊宏譯

出版年：2013

出版社：天下遠見（臺北市）

頁 數：318 頁

越來越多科學家肯定，在大數據 (Big Data) 時代，即使不是醫藥專業人員或防疫人員，也能幫助人類控制疫情。權威醫療雜誌刺絡針《The Lancet》報導，根據國際航空運輸協會提供近兩年伊波拉疫區所有飛航資料，精確計算出西非疫區出來的班機最熱門的航點，是中國大陸的北京、上海、廣州，其他是香港、印度孟買。該數據顯示防疫網的優先順序，指出防疫風險的第一線，而沒有科學根據的防疫對策，不但浪費資源，亦可能惡化問題。因此，伊波拉的全球防疫戰線能精準上路，關鍵在開放數據 (open data)，透過資料的透明化，讓全球菁英共同尋求對策，有效與即時的控制疫情散播。巨量資料被產業界視為第五波科技革命後的明星

* 陳啟清為開南大學公共事務管理學系副教授。

Chii-Ching Chen, Associate Professor of Department of Public Affairs and Management, Kainan University.

產業，人類社會將出現以資料為驅動力的經濟（data-driven economy），改變現有市場、組織、公民與政府之間的關係，亦為商業、市場和社會帶來變化。當全球都在大數據中淘寶，臺灣政府與企業部門亦應正視此一趨勢。為此，本篇針對麥爾荀柏格（Viktor Mayer-Schonberger）、庫基耶（Kenneth Cukier）合著，林俊宏翻譯，由天下文化公司出版的《大數據》提出書評。

在《大數據》一書中，作者以系統與務實的方式呈現巨量資料革命的趨勢與衝擊，提供我們面對巨量資料的重要性與發展。作者在書中所舉的例子都非常務實，應用議題從醫學研究、地下水道、交通運輸、犯罪預防到國家安全，參與組織從企業到各層級的政府，值得參考。隨著巨量資料的預測越來越準確，許多國家已經開始採用這種大數據模擬出來的資料做為執法單位的決策參考。例如美國衛生人力部利用大數據分析過去行政訴訟的資料，藉以改善其決策品質。西班牙馬德里政府，整合警察、消防、醫療系統，使救援時間大幅度縮短，巡邏隊、消防車、救護車能夠在 8 分鐘內到達 81% 的突發事件現場；在新加坡，智能交通綜合資訊管理平台在預測交通流速和流量方面有高達 85% 的正確率，透過有效的引導和介入，提升尖峰時段的車輛通行效率。

《大數據》全書分為十個章節，開頭前四章先提出概括式的介紹，讓剛接觸巨量資料概念的新手可以熱身一下，從大數據的意涵、範圍與特質的了解起。大數據簡稱做巨量資料或是海量資料，指的是所涉及的資料量規模巨大到無法透過目前主流軟體工具，在合理時間內達到擷取、管理、整理，成為幫助企業經營決策更積極目的的資訊。一般而言，巨量資料的特質常以四個 V 描述，包含：數據量巨大（volume）、數據多樣性（variety）、處理速度快（velocity）、資料的真實性（veracity）。資料的真正價值，就像是漂浮在海上的冰山，第一眼看到的只有小部分，有很大一塊都藏在海平面之下。如果組織能看到此項特點，就能獲取隱藏價值，獲得巨大利益。此外，有些時候，想要釋放潛藏的資料價值，就必須與其他資料結合，甚至是和截然不同的資料結合。因此，靠著用全新方式混合資料，就能達到創新。

在《大數據》一書的第五、六章說明巨量資料的價值、優勢與形成的環節。巨量資料的提出並不是要「訓練」電腦如何像人類一樣「思考」，而是透過計算大量的資料，以此推斷機率，作為效率決策的基礎。例如某封電子郵件是垃圾郵件的可

能性；輸入的「teh」應該是「the」的可能性。尤其巨量資料的內容常常是混亂不齊、品質不一，分布在世界各地的無數伺服器中，因而顛覆以往認為「數大即是美」的觀念。在大數據時代，資料多、未必周全，其可能代表雜亂無章，但卻能「亂中有序」。因此，處理巨量資料的時候，只要能大概有個方向即可，不需要講究到一寸、一分。而在微觀層面上失去的精確，可以用宏觀層面上獲得的觀點及見解加以彌補，因此，大數據改變我們將「精確」、「一致」奉為主臬的思維。

在分析巨量資料的特質與價值後，作者在第八與九章分別針對巨量資料的風險、缺點和管控方式，說明巨量資料在應用面的問題。首先，大數據的運用涉及非常敏感的個資隱私問題，個資隱私問題不僅發生於企業對於民眾個資的蒐集、利用與處理，亦發生於政府部門對於私部門與民眾資料的蒐集與利用，其可能侵犯隱私、降低人民自由與增加國家與企業公司的控制。史諾登（Edward Joseph Snowden）的事件彰顯一項事實：即便是民主國家的美國政府，仍免不了濫用個資保護的法制針對人民進行秘密資料的蒐集。顯示在大數據時代，每個人在不知不覺中「被蒐集」，再加以資料整合與加總。換言之，大數據可以創造一個人的所有背景資料（包括本人可能都不知道的資料），儼然是現代社會的「老大哥」（Big brother）。其次，作者提出在使用巨量資料可能面對另一種獨特且令人不安的問題。有了大數據之後，組織對於個人的判斷，可能不是依據他們的實際行為，而是巨量資料顯示的預測。例如美國聯邦政府假釋委員會已經依據此一概念，建立「假釋人員再犯預測治安系統」。根據假釋犯所有過去的犯罪模式，及他在監期間的一些行為個別資料，再將假釋犯過去接觸過的環境，以及假釋以後可能去的環境，其中一些與假釋犯犯罪因子的各種參數作模擬比對，進而估算其再犯的機率，並將前述資料分析的結果作為某個囚犯是否能夠假釋的參考資料。換言之，大數據顯示的預測資訊，超越個人的實際作為，稍微誇張的說法是各種數據的加總可以拚湊、製造出「一個人」，因而數據對於組織決策做為是商機或利基，但對於個人可能是「被透明化」與「被攤在陽光下」的危機。

因此，作者在歌頌大數據的同時，於《大數據》第十章指出大數據由量變而產生的質變，確實為人類未來的生活帶來許多的便利性與效率性，同時對於組織或社區，亦促進正向的發展。但是一如「科技來自於人性」，「數據始終來自於人類活動、回饋於人類活動」，作者最後提醒在關注大數據的同時，仍要回歸對於「人」

的關懷，提醒讀者要謙卑、要注意道德及人性化的一面。由此可知，作者在宣揚大數據的趨勢與優勢之後，最後透過對於人性關懷的呼籲，沉澱大數據對於人類社會的衝擊，提醒人類社會不要被大數據的浪潮所迷失與淹沒，能夠回到大數據被提出的初衷：幫助人類社會更有智慧的解決問題，而不是更便捷的製造問題。

讀者透過作者有系統的概念與實務的說明，在閱讀《大數據》這本書後，的確能體會到，在二十世紀，價值從過去的有形資產（例如土地和產房），轉移到無形資產（例如品牌和智慧財產），現在更進一步擴展，使「資料」成為重要的企業資產、必要的經濟資源投入，以及新商業模式的基礎。同時認知到大數據可以存在於各個領域，包括醫療健康、食品安全、氣候、能源與環保、智慧交通等，在不違反個資隱私保護的原則下，如何妥善利用這些巨量資料，挖掘有價值的知識與決策資訊，成為當前政府必須面對的一項大挑戰。

尤其特別值得關注的是，對於前瞻趨勢發展有興趣的人，即使不是資訊科技或是資訊工程的專業背景，亦能從《大數據》這本書中體會資訊科技對於人類社會進步的影響與衝擊，尤其顛覆以往認為「數大即是美」的觀念，改變過去「先建立假設，再蒐集資料印證」的求證方式，以及不同類型與領域資料的「混搭」，亦即面對巨量資料時代的來臨，先要徹底改變以往小數據時代的思維。

政府作為各種客觀（業務申辦、會議紀錄等）與主觀數據（民意信箱、陳情等）的最大擁有者，亦必須理解大數據透過科技的創新與發展，及數據的全面感知、蒐集、分析與共用，亦為政府跨部門與跨資訊系統汲取、分析、管理以及分配資料的方式帶來挑戰，促成政府決策行為應根據事實與數據的分析判斷。大數據對於政府部門的影響包括：第一，促進資料的分享：由於政府機關的數量與性質龐雜，各自所擁有的運作知識或資訊相當豐富。透過大數據的平台，可以有效地協助政府機關分享彼此的資訊，藉以提升其解決問題的能力。第二，促進組織的學習：管理階層可以透過大數據的分析，發掘單位在管理過程中未發現的可能問題或重要關鍵，針對調整可能的偏差方向。第三，促進行政事務的品質：透過大數據分析預測特定民眾的特質，可以據以針對其個人的特質與需求進行客製化的調整。第四，促進問題的解決：大數據發掘複雜資料隱藏的趨勢價值，據以提出相對應的解決方案，並研擬可能的因應對策。最後，促進決策的創新：利用大數據分析可以發掘以往政府運作浪費或低效率的部分，透過資源重新分配的方式，達到整體績效最大化的程度。

以往討論政府之間的跨域治理時，多以組織合作與業務協調建立跨域互動模式，例如透過跨機關的委員會方式，討論與協調共同行動，換言之，人員之間的觀念釐清與共識建立，扮演跨域治理的重要角色，惟常常「成也在人、敗也在人」，甚至發生「蕭規曹不隨」的情形，使得跨域合作充滿模糊性與不確定性。隨著大數據時代的來臨，透過各機關擁有的多元與大量的數據，其對於跨域治理亦產生加持作用，例如以資料共享與預測分析方式，對於地方政府共同面對問題，透過數據評量主觀需求與客觀條件的可能組合，作為討論共同行動與成效預測的基礎。因此，在臺灣，政府有責任扮演開放數據的領頭羊，甚至從跨域治理領域的決策行為開始運用，透過大數據解決以往跨域互動時需要協調的問題，讓大數據呈現的現況與趨勢為跨域協調議題「說話」。而有鑑於美國估計到二〇一八年，需要二百萬具有大數據知識與專業能力的人才，因而，臺灣政府先從教育層面吸引與訓練大數據相關工作的人員，尤其是增加數據科學家（data scientists）的職位；以及推動企業組織對於具有數學、統計等相關領域的人員進行內部訓練，使其成為數據科學家。

最後，《大數據》這本書作者對於大數據的意涵、特質、應用與限制提出系統性與全面性的說明，惟作者提出的是他國的應用案例，在國情與文化的差異下，對於實務的理解與應用有「隔靴搔癢」之憾。有鑑於臺灣不能自外於全球大數據的浪潮，期許未來亦能有臺灣本土化的大數據實務應用專書的出版。