

大數據是雙面刃，善用不擅用

廖怡茜*

壹、前言：雜亂痕跡成為數據來源



圖 1 中央流行疫情指揮中心於 2 月 7 日發布之緊急警報簡訊。

2 月 7 日下午 5 時 45 分，讓人惴惴不安的警示聲此起彼落，路上熙來攘往的人們不約而同地拿起手機，小小方框中顯目的三角形驚嘆號警示訊息包含著一個連結，連結至鑽石公主號 1 月 31 日停泊基隆港後，該船旅客下船後造訪的路線軌跡，提醒曾於該日赴基隆廟口、九份、臺北 101、西門町等北北基地區重要景點的民眾自主健康管理。

*法務部調查局調查官。

該警訊通知係中央疫情指揮中心跨部會協調，向國內電信業者調閱 1 月 31 日從基隆港周邊到旅客移動時的手機基地台訊號資料，分析數據後，回溯勾勒鑽石公主號旅客下船後的移動地圖（圖 2），再透過「災防告警細胞廣播傳染病警示訊息發送系統」（Public Warning System, PWS）發布「國家級警報」。誠如《數據、謊言與真相》作者賽斯·史蒂芬斯—大衛德維茲（Seth Stephens-Davidowitz）所言：「我們在生活中留下的雜亂痕跡，正成為數據的主要來源」，一個又一個人們不經意留下的基地台訊號資料，看似雜亂的生活痕跡，成為拼湊旅客移動路徑的電信數據來源，描繪人們未能如實記憶，卻倍顯重要的事實軌跡。然而，擁有大數據並不會產生精闢見解，仍需要人類調查、分析與判斷。

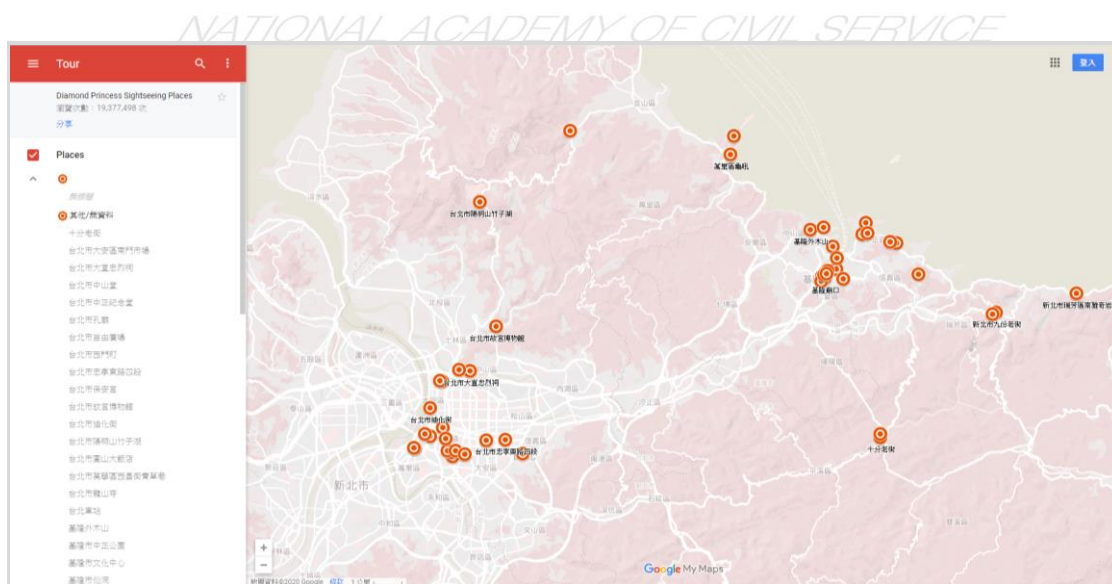


圖 2 中央流行疫情指揮中心透過基地台等資料勾勒製作之鑽石公主號旅客行程地圖

貳、大數據：靜靜地與你我的心對話

一、何謂大數據？

何謂大數據？透過 Google 檢索「大數據」、「定義」，瀏覽相關文章，歸納目前對於大數據特性的普遍定義為 4V，包括資料量（Volume）、資料傳輸速度（Velocity）、資料類型（Variety）及真實性（Veracity），具有此等特性之資料被稱為大數據。然而，翻遍《數據、謊言與真相》一書，找尋作者如何定義「大數據」，卻遍尋不著，但作者提到「字詞是數據，點擊是數據，連結是數據，拼寫錯誤是數據，夢到香蕉是數據，語調是數據，喘息聲是數據，心跳是數據，脾臟大小是數據」、「超市結帳隊伍的照片是有價值的數據，超市商品數量充足與否是數據，蘋果成熟度是數據，從外太空拍攝的照片是數據，嘴唇彎曲度是數據，一切都是數據」，並言「人們搜尋資訊，這種行為本身就是資訊」，筆者剛剛 Google 檢索關鍵詞「大數據」、「定義」這個行為，亦化為「大數據」，驗證著「我們在生活中留下的雜亂痕跡，正成為數據的主要來源」。

二、大數據的四種力量

受惠於科技進展，資料取得與儲存成本均大幅下降，造就大數據興起，讓過往未具如此龐大價值的雜亂痕跡，蛻變為擁有四種力量的大數據：

- (一) 提供新類型數據：各種議題均有獨特數據來源，並仰賴電腦運算處理成千上萬數據，讓使用者能科學剖析過往僅能臆測的領域，做出有意義分析。
- (二) 提供誠實數據：此係仰賴「以為沒人知道」的力量，人們面對搜尋引擎時猶如吃了「數位誠實豆沙包」，曝光人們深藏內心的秘密，瞭解人們真正想要與做了什麼，而非人們說自己要什麼與做了什麼。
- (三) 針對更特定群體進行分析：大數據讓獲取每個樣本的成本大幅降低，反映大趨勢，並進一步針對特定群體進行分析，允許分析者可以有意義放大檢視數據集的細部，取得新見解。
- (四) 更易進行「因果關係」實驗：隨機實驗 (A/B Testing) 係將實驗對象分為實驗組與對照組，大數據讓隨機對照實驗變得更容易進行。

三、大數據無所不在

雖未明確定義「大數據」，但作者透過大數據分析案例論述大數據的四種力量，並在《數據、謊言與真相》第三章「怎樣的數據算是大數據」中結論「一切都是數據！利用這些新型數據，我們終於可以揭穿人們的謊言」。無論係網頁搜尋紀錄、網站停留觀看時間、信用卡消費紀錄、手

機基地台訊號等等，大數據讓世人真面目無所遁形，與其推崇或讚賞，與其害怕或抗拒，與其理性定義大數據，不如言，生活裡，大數據無所不在，走出書房，走進天地、百姓間，周遭處處充滿大數據的身影，它，靜靜地與歲月、與你、我的心對話。

參、大數據還是大錯誤

大數據存在於各個領域，涵納醫療健康、食品安全、氣候氣象、能源與環保、智慧交通、行銷推廣等等，無論私部門或公部門，利用目前可取得的所有新型數據，思考需要何種數據、何者可做為數據使用，經過資料分析，即可萃取價值。然而，大數據係萬能的嗎？大數據是巨量、海量，但卻非常細微的數據聚合，大數據不是一批猶如土石流般傾瀉而下的數據，而係一場沙塵暴。在沙漠裡，陣陣強風捲起無數沙粒，可能蒙蔽人們雙眼，讓人失去方向，迷失在無垠荒漠。

一、維度的詛咒：勿拿大數據來算明牌

「維度的詛咒」係指變數很多，但觀察數不足，恐讓某個變數成為「幸運者」，讓人誤以為此「幸運者」是很好的預測變數，惟實係巧合、偶然而已。作者舉例，假設每天擲一千枚硬幣，記錄投擲結果，並記下當天標普指數上漲或下跌情形，其中，編號 391 號硬幣出現正面時，標

普指數上漲機率達 70.3%，以統計方面來說，兩者高度相關，因此每天擲編號 391 號硬幣並擲出正面，即可買進股票。倘若以此預測股市漲跌，便落入「維度的詛咒」。

二、大數據：新的階級落差

在資訊時代，資訊本身即是一種資本，擁有資訊與否成為決定人們能否獲益的重要因素，資訊科技的近用機會與使用經驗會因個人的性別、種族、階級或居住地理區域等社經背景不同而有所差異，形成「數位落差」(digital divided)，而「大數據」(big data) 是否亦製造組織、機構、個人間的新落差？大數據興起係因數據每天巨量產生，社群內大量訊息交換，資料存量與流量都成指數成長，同時因資料存取技術進步，成本大幅下降，因此擁有資料庫者，可藉資料平台間交叉分析與比對，獲得更多新情資以協助組織、機構、個人等進行判斷。因此數據的擁有 (have) 和未擁有 (have not) 恐形成新的階級落差。

三、大數據：監控巨靈

《數據、謊言與真相》封面揭示「人會說謊，但大數據不會」，大數據揭露人們的真面目，掌握大數據的人、組織、機構可以獲悉更多資訊，但無法取得這些數據的人呢？大數據係研究人類行為的利器，同時製造極大資訊落差，且大數據的力量之一來自於「誠實」，但人們並非自願誠

實，而係「以為沒人知道」。

(一) 科技、防疫、大數據

2019 新型冠狀病毒 (COVID-19) 肆虐全球，中央疫情指揮中心 2 月 20 日公布感染新冠肺炎過世的中部白牌計程車司機感染源係一名浙江臺商。當日，行政院副院長陳其邁以「科技、防疫、大數據：政府如何查出第 19 例白牌車司機的感染源？」為題 (圖 3)，在臉書揭密政府找出感染源的方法就是透過「大數據」及跨部會全力合作，動用警政、民政等多個系統，再加上監視器、通聯紀錄等進行疫情調查與分析，俾利確診個案回憶曾從事的活動，以便政府掌握個案曾出入的場所、接觸史，篩選疑似對象，圍堵防疫可能的破口。

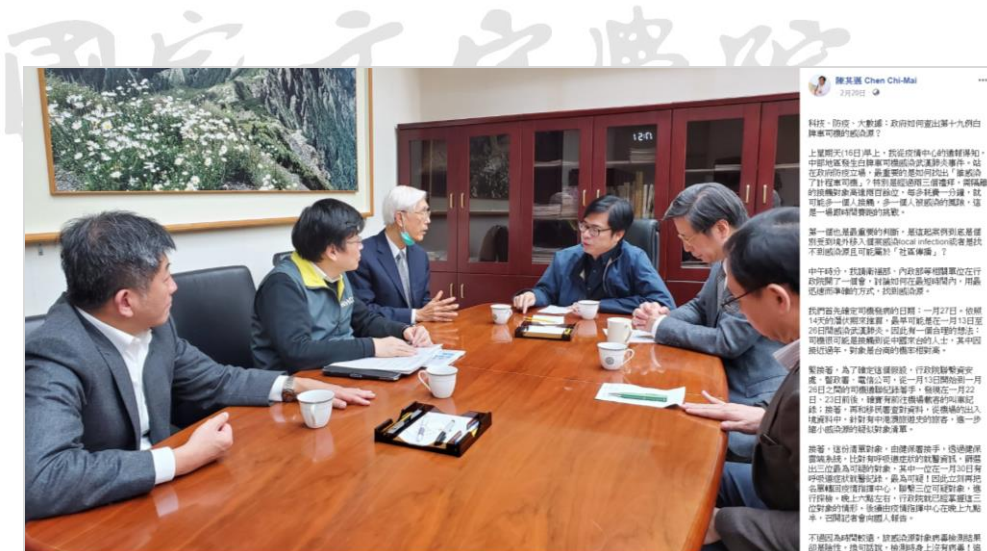


圖 3 行政院副院長陳其邁以「科技、防疫、大數據：政府如何查出第 19 例白牌車司機的感染源？」為題，於臉書透露政府追查感染源之情形

(二) 安靜的恐怖主義

政府嚴防 COVID-19 疫情，獲得國人肯定，亦成為國際輿論稱讚焦點，詎料此次防疫採取的跨部會資料串接、電信監控、口罩實名制等措施，卻遭到不當使用。以健保資料勾稽出入境資料為例，個人旅遊史揭露，固然有助於穩固防疫網絡，但隨著全球疫情升溫，資料揭露範圍擴大，開放無卡化查詢旅遊史服務後，臺北市醫師職業工會在臉書發文控訴，醫院單位主管濫查員工出國旅遊史的不合理作為（圖 4）。在《大數據的傲慢與偏見》一書中，作者凱西·歐尼爾（Cathy O’Neil）稱不當的數據運用為「安靜的恐怖主義」，而此或正應證著作者所言。

NATIONAL ACADEMY OF CIVIL SERVICE



圖 4 臺北市醫師職業工會 3 月 4 日在工會臉書專頁發文控訴醫療院所濫查員工旅遊史的不合理作為

(三) 數位圓形監獄儼然成形

大數據窺探人們自以為獨處的時刻，記錄人們生活中的雜亂痕跡，無論係搜尋引擎、社群媒體、行動通訊訊號、照片、文字、影音等等，並將各類「數位足跡」積累成網路大數據，大數據時代的「圓形監獄」儼然成形。大數據為維護治安、國家安全、醫療、商業及其他重要領域帶來無邊利益，卻同時衝擊個人隱私。歸納大數據可能對社會造成如下衝擊：

1. 弱勢者成為剝削目標

在《大數據的傲慢與偏見》一書中，作者以美國鳳凰城大學此類被稱為「文憑工廠」或「學店」的學校為例，該等營利大學於 2004 至 2014 年間，入學人數增加兩倍，該些學校如今占美國的大學學生人口 11%，原因即在網路廣告。該書作者指稱，鳳凰城大學花在 Google 廣告的支出，至今已超過 5,000 萬美元，該集團花在每位學生身上的行銷費為 2,225 美元，但真正用在教學支出卻只有 892 美元。營利大學以脆弱的低下階層為目標市場，以向上流動為誘餌，致力吸引窮人就讀，甚至捏造學生就業率、濫收學雜費等惡行。

2. 誤用或濫用個人資訊

例如，中國大陸透過當地網路企業的資訊，包括消費、行為、人際關係、身分特質與言論、履約能力與信用歷史等等，運用大數據監控推出全面性「社會信用制度」，當社會信用評等低的時候，甚至無法購買機票或高鐵，而此，可能造成人民為取得良好信用，購買國家認同的產品、對政治不正確議題噤聲等等。

3. 歧視

倉儲管理與定價決策皆已採自動化的全球知名跨國電子商務企業亞馬遜公司（Amazon）於 2014 年開始建置試驗性徵才人工智慧（AI）程式，原預期可協助檢閱求職者履歷，宛如亞馬遜消費者對商品的評價，一到五星擇一，藉此初步篩選出百份履歷中的前五強。惟，研發人員隔年發現，在軟體開發等技術相關的職務上，AI 徵才並未秉持性別中立原則，原因係 AI 運用過去 10 年來的履歷資料與招聘數據，但當初求職者多為男性，造就今日科技業的男人天下；這個現實讓 AI 誤以為男性求職者較佳，凡有「女性」字眼的履歷均打低分，即使修正程式仍難保不再歧視，致使亞馬遜於 2017 年初正式宣告放棄此計畫。

4. 職場身心傷害

承上案例，亞馬遜企圖用大數據與 AI 篩選履歷及招聘人選，最終因大數據的性別偏誤而終止，惟亞馬遜又遭揭露以數位追蹤器記錄物流倉儲員工效率，追蹤員工摸魚時間，並由系統自動解僱未達標準的員工，從 2017 年 8 月至 2018 年 10 月，巴爾的摩倉庫即有 300 位員工因此被解僱，占該倉庫雇用人數的 12%，而該措施致使員工在高壓且精密監視下，無法好好上廁所與喝水。此類運用大數據進行的科技監控，造成被監控者壓力，影響身心健康。

5. 政治操作與社會分離

利用假訊息進行資訊攻擊並透過媒體操作輿論、分化社會群體，達到特定政治目的的作法已受到各國重視，惟運用大數據的演算法強化網路同溫層效應與社群網站的過濾機制，致使網民接觸到更多單一面向、渠等想要看到的內容，而非多元訊息與意見，加速錯誤與仇恨訊息散布，造成社會分離。

6. 數據與系統偏誤

如亞馬遜透過 AI 大數據篩選履歷一案，由於取樣偏誤或在設計程式時，已將個人偏見在無意識情況下寫進程式中，造成

分析結果偏誤。

肆、用數據讓公共治理更有力量

大數據猶如雙面刃，有其邪惡與偏誤，有其良善與效益，若數位科技與大數據能被更具公益目的運用，勢將有助於讓公共或企業服務運送至各地並擴及各階層，協助弱勢群體發展、解決氣候變遷、增益能源使用、強化環境保護等等。

運用大數據可讓政府在政策問題形成、政策議程設定、政策規劃等過程更精準掌握資訊，政策執行後，藉由大數據分析，快速瞭解執行成效與反饋，即時回應民眾需求，擘劃更完整的政策選案。例如，為防止新冠肺炎疫情，健保署提供「特約藥局與成人、兒童口罩庫存數量查詢系統開放資料」，零時政府 g0v.tw 在共筆平台 HackMD 整理口罩供需資訊平台，鼓勵各界上傳自己開發的服務，包含政府官方系統在內，多種地圖服務讓瀏覽器定位使用者位置，便能自動搜尋周圍健保藥局與剩餘口罩數量，便利民眾查詢與前往購買，且政府藉大數據瞭解各地區供需狀況，適時調配口罩配送數量。又或者，透過網路輿情分析，讓相關部會在某項議題爆發時可以迅速得知網路輿論與民眾反應，或是運用網路及社群媒體等多樣化管道論述政策，將政策更完整呈現，提高政府的回應速度與透明度，並以良好的政策行銷來落實政策目標。

伍、結語：大數據是一種精神，善用不擅用

一、大數據不僅是資源，亦係武器

大數據的發展與運用隨著人類的「想要」與「需要」而生，有其能與不能，近年數據與資訊科技進展，從交通工具、家電、穿戴式裝置、手機與網路活動、散佈在各街道巷口、各角落的監視器與感測器，記錄人們生活的一舉一動，生成大量數據，並用於產生各種預測、研究、商業價值，亦係人工智慧與訓練機器學習的主要資源。大數據不僅是資源，亦係武器，掌握數據並具有運用這些數據能力者將更有機會主導經濟與社會發展。大數據可能變成當權者控制社會、組織高層監控員工的工具，亦可勾勒鑽石公主號旅客移動路徑、白牌車司機接觸史、即時口罩庫存地圖等等，強化政府效能，精進公共治理。

二、大數據與傳統作法需相輔相成

觀察到「現象」，未必等於掌握「真相」，亦不等於理解「原因」，研究人類行為與社會現象，質性觀察、歸納與演繹仍不可或缺，爰此，作者在書末預測，「大數據所揭示之物，將引發一場革命，但這並不代表我們可以向數據丟出任何問題，大數據也無法將人類數千年來為瞭解世界而發展出的所有方式去除，大數據與傳統作法，其實相輔相成」。或言，

大數據，是一種精神，是研究人類與社會的嶄新思維，奠基在傳統上的創新。在不違反個資隱私保護前提下，妥善利用大數據，挖掘有價值的決策資訊並形成決策知識，完善公共治理，係政府須有的決策與治理思維。今日，公、私部門不斷透過效能更佳的技术分析大數據，更深入理解人類行為的內在因素和外顯特徵脈絡，協助預測未來發展趨勢，據以做成政治決斷或商業決策，精進公共或企業治理。大數據是雙面刃，若濫用大數據，終將破壞民主與人權，讓社會風險調控失靈，產生深鉅後遺症。大數據本身是中性的，端視使用者善用或擅用！

NATIONAL ACADEMY OF CIVIL SERVICE

國家文官學院

參考文獻

Cortilia Lin. (2018)。大數據的迷思一：平台不等於數據分析。哈佛商業評論。

檢自 https://www.hbrtaiwan.com/article_content_AR0008151.html

高振源 (2019)。近年我國政府大數據政策與經費執行之探討。國會季刊，1，64-89。

齊立文 (2018)。大數據一點都不大？可能誤導你的決策？破解 Big Data 的 8 個迷思。經理人。檢自 <https://www.managertoday.com.tw/articles/view/56401?page=1>

林昕璇 (2018)。從史諾登到臉書個資外洩，大數據時代的隱私困境與規範難題。檢自 <https://www.thenewslens.com/article/96901>

余至浩 (2017)。數位政府決策要有新思維，讓數據帶頭說話。檢自 <https://www.ithome.com.tw/news/117276>

黃東益 (2017)。資訊通訊科技驅動治理轉型？趨勢與研究議題。文官制度季刊，3，1-25。

羅鈺珊 (2018)。數據經濟下共融成長的挑戰：大數據的兩面刃。經濟前瞻，178，87-93。

凱西·歐尼爾 (2017)。大數據的傲慢與偏見：一個「圈內數學家」對演算法霸權的警告與揭發。臺北市：大寫出版。