登泰山而小天下

—從雲端看臺灣，《大數據》給我的啓示

【佳作·馮佩棻】

壹、前言：揭開巨量資料的神祕面紗

1914 年 6 月 28 日，奧匈帝國王儲在賽拉耶佛遇刺，吹響了第 1 次世界大戰的號角。歷史學家估計這場持續了 5 年的戰爭，造成 1 千萬人口喪生。不同於以原子彈畫下驚歎號結束的第 2 次世界大戰，一戰終止的原因，並不是任一陣線以武力征服敵國獲勝，而是在戰爭末期席捲全球的西班牙流感。當時年輕力壯的士兵與平民相繼因流感或其引起的併發症死亡，各國皆無法投注充足兵力參戰，就連談判桌上的美、英、法元首也罹病。最終因流感喪生的人數高達數千萬，遠大於直接因戰爭而罹難的人口，遑論人民因疾病蒙受的經濟損失。

2009 年 3 月，墨西哥及美國陸續出現罹患不明原因急性呼吸道疾病的個案，經防疫人員採集檢體鑑定，直到同年 4 月中旬才確認感染源來自基因重組的新型流感 H1N1 病毒。世界衛生組織（WHO）於同年 4 月 24 日的公告指出，美墨兩國新型流感疫情延燒已造成社區型流行，在墨西哥首都墨西哥城有 854 人產生流感併發症，其中 59 人因此死亡。2009 年的新型流感與 1918 年的西班牙流感病毒株分型同樣是 H1N1，所幸 2009 年新型流感病毒致死率較低，藉助流感疫苗施打與抗病毒藥物治療雙管齊下防堵，人類總算打贏 21 世紀第 1 場流感大流行的戰役。

巧合的是，早在新型流感爆發前 1 個月，搜尋引擎巨擘谷歌（Google）與美國疾病管制與預防中心的研究人員在國際科學權威期刊《自然》上

共同發表文章，宣告「谷歌流感趨勢」（Google Flu Trends，簡稱 GFT）的誕生。研究人員比對 2003 年至 2008 年間美國最常被搜尋的 5 千萬個詞語與同期間官方流感監控資料，找出 45 個與類流感盛行率相關性最高的關鍵字。藉著分析各地區關鍵字被搜尋的次數，GFT 可即時預測出每天美國各州的流感疫情發展。由於 GFT 能掌握民眾自覺發病至就醫確診通報前的時間落差，故擁有比官方資料更迅速的時效性，相較於美國首例 H1N1 個案從發病至確診整整花了 2 週時間，GFT 的效率可說是讓傳統疫情通報系統望塵莫及。

乍看《大數據》的書名可能會覺得本書很艱澀，但經由作者生動介紹各領域運用巨量資料的實例一舉凡「谷歌流感趨勢」、機票票價預測網站「Farecast」與預測飛機到達時刻網站「FlyOnTime.us」、亞馬遜書店的推薦書單等，我們能深入淺出的了解巨量資料對日常生活的影響。藉著此次專書閱讀心得寫作的機會，筆者從自己最熟悉的防疫工作起步，跳脫蕭規曹隨的窠臼，從另一種宏觀的角度思考該如何在實務上運用巨量資料理論。身處在資訊大爆炸的時代，資料蒐集與處理資料的進展一日千里，巨量資料分析將帶領我們從見樹轉為見林的宏觀視角，整合各領域資源，突破傳統技術空間與時間限制，洞燭機先掌握時代脈動，提供民眾更便利的服務。

貳、重點歸納：我看《大數據》

本書分為 10 章，除了第 1 章簡介「現在」與第 10 章結語「未來」之外，第 2 至 4 章介紹巨量資料的三大特性—「樣本=母群體、雜亂、相關性」；第 5 至 9 章則是以實例說明我們該如何取得巨量資料，發掘巨量資料的價值，並管控巨量資料帶來的風險。茲將書中內容歸納為「巨量資料的三大特色」與「巨量資料應用三部曲」，分述如下：

一、巨量資料的三大特色

（一）樣本＝母群體

以往因為用來蒐集、組織、儲存和分析資料的工具受

限，大多時候我們只能蒐集全部資料的一小部份，猶如以蠡測海，以管窺天。中央極限定理是古典統計學的基石，該定理指出：「無論觀測變數母群體呈何種機率分配，只要抽樣樣本數量夠大的時候，樣本平均數的抽樣分配會趨近於常態分布」，因此研究者可透過計算抽樣的樣本平均數推測母群體真實狀態。然而中央極限定理須符合樣本隨機抽樣的前提，一旦蒐集資料的方式有系統性誤差，就會影響結果信度。巨量資料放大了樣本數的規模，使研究者不再透過繁複的抽樣手續，而是直接與整個母群體面對面。作者舉國際匯款 Xoom 公司為例，該公司利用巨量資料技術逐一檢查每筆交易，讓原本隱蔽在樣本平均值中魔鬼的細節現身，找出來自犯罪集團的異常資料，進而有效防堵信用卡詐騙案件發生。

（二）雜亂—不精確也無妨，數大便是美

過去我們深信藉由測量某個現象，我們就能理解這個現象。在抽樣的世界裡，對精確的要求極為重要。因為只分析少量資料點的情形下，錯誤可能放大，也就會減少整體結果的準確度。當我們可以取得樣本數近似母群體的資料時，我們就可以忽略資料細節的雜亂，轉而巨觀完整的大趨勢。就像是印象派繪畫，近看只是凌亂的筆觸，但向後退一步看，就是偉大的畫作。就連向來將精確奉為圭臬的資料庫設計，為了廣納整合不同來源資料，也開始吹起接受雜亂的風潮，捨棄絕對精確的堅持，甚至允許使用者自訂標籤，以求可以提升處理效率，這正是「泰山不讓土壤，故能成其大；河海不擇細流，故能就其深」的精神。

（三）相關性至上

追根究底是人類的天性，也是科學進步的動力。正因為牛頓去探求蘋果落地的原因，才發現了萬有引力，進而推導出力學三大定律，引發 18 世紀工業革命。然而，對於末端

使用者跟企業來說，原因並不是最重要的事，只要知道大致可能的趨勢，就可以做出決定或是行銷策略。巨量資料讓相關性應用更加簡單、迅速，甚至不用預先設立假說，就能同時以多種數學模型進行多種變因分析。亞馬遜書店、沃爾瑪量販店與美國折扣零售商塔吉特等大企業早已搶得先機，運用消費者購買紀錄的巨量資料預估個人消費偏好，推出客制化行銷策略，大幅提升獲利。

二、巨量資料應用三部曲

（一）資料化

資料化係指將想探討的現象以量化呈現，以便整理分析。資料的範疇不限於電腦可以判讀的資訊（0 與 1 的組合），透過現代數位測量與儲存裝備，無論是文字、圖片、影片、音樂、使用者所在地、甚至是社群網站的互動資料，皆可輕易完成數位化。然而數位化並不等於資料化，以谷歌建立線上文本資料庫為例，谷歌先將文本以掃描存檔數位化之後，再利用光學文字辨識軟體分析其中的字母、詞彙、句子和段落，完成資料化。如此一來使用者不但可以輕易檢索文本的內容，更能分析比對文本間資料，追溯特定概念的起源。谷歌也藉由比較各種語言譯本用語，修正其翻譯軟體功能，創造更多附加價值。

（二）發掘價值

資料的真正價值，就像是漂浮在海上的冰山，第一眼看到的只有一小部分，還有更大的價值藏在海面下等待創新者發掘。資料可謂資訊經濟時代的石油，作者提出「重複使用資料、重新組合資料、讓資料買一送一」三項法則，指引我們如何從既有資料中發掘出新價值。近來公部門也開始正視資訊公開帶來的效益，美國在歐巴馬總統大力推動下，建立了政府資訊公開平臺 data.gov，截至目前（2014 年 7 月 1 日）

已累積跨 229 個部會機構，超過 11 萬筆公開資料集。2013 年 5 月 9 日，歐巴馬總統再度發出命令要求政府部門公布所有可對外公開的資料，並轉換為機器可分析的格式，又進一步增加民間應用公開資訊的可近性。我國亦於民國 101 年 11 月 8 日第 3322 次行政院院會決議，成立政府資訊公開平臺 data.gov.tw，截至目前（2014 年 7 月 1 日）為止共有 2017 筆資料集開放，各部會與各級政府也將陸續加入資訊公開的行列。

本書作者認為資料價值鏈由資料所有人、資料專家及具有巨量資料思維者組成（圖 1）：資料所有人提供原始資料給具有運算、分析能力的資料專家，資料專家可以設計數學模型了解資料與主題的相關性，但最重要的創新價值是掌握在能洞見機會，提出創意主題的巨量資料思維者。隨著巨量資料成為顯學，跨組織的資料中介機構也應運而生。中介機構說服資料所有人分享手中的資料，在不影響所有人利益的情況下，讓所有分享者均霑整合巨量資料帶來的創新價值。

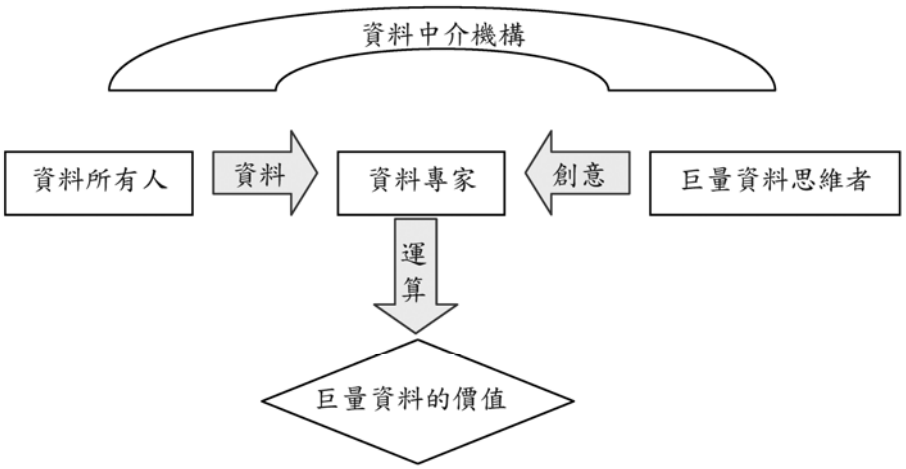


圖 1　資料價值鏈的架構
資料來源：筆者自繪。

（三）風險管控

蒐集巨量資料的工具越聰明，意味著對我們生活帶來越多監控。因為無法預測個人資料未來被使用的用途，我們無法仰賴個人資料保護「知情同意」與「匿名化」的原則約束使用巨量資料的企業或政府。更甚者，企業或政府濫用巨量資料的預測，將相關性誤解讀為因果關係，抹煞了個人良知抉擇和自由意志，隨意將他人貼上標籤，自以為以巨量資料實踐正義，其實是被禁錮在機率的牢籠中，成為資料的俘虜。

為了追求報導速度並引起閱聽大眾關注，近來媒體充斥著未經查證逕行以小道消息、爆料發布的「迭代新聞」。當新聞被證實為錯誤消息，媒體往往不會再次發布修正的新聞，而是在原有的新聞下方加註更正啟示。一旦錯誤的新聞被廣泛引用討論，加上搜尋引擎頁庫存檔推波助瀾，更正啟示非但無法還原真相，反而加深讀者錯誤印象，讓迭代新聞的受害者背負難以抹滅的紅字。為了保障人民的「被遺忘權」，歐盟最高法院於 2014 年 5 月 13 日宣判，要求谷歌需移除該公司搜尋引擎上涉及個人敏感、無關、不當或過時資訊。谷歌於同年 5 月 30 日起受理歐盟民眾申請，首日即湧入 1 萬 2 千筆以上案件，超乎該公司預期。本書作者也提出應依照個資風險設立不同使用時限的觀點，督促資料使用者儘快於設定用途內使用個資，如逾時效就必須將資料刪除。

本書作者提出三大策略作為管理巨量資料風險的準則：

1. 將個資保護由「個人同意制」，轉成資料使用者的「使用責任制」。
2. 運用巨量資料進行預測時，要尊重每一個人的能動性。
3. 培養巨量資料的審計師—「演算學家」。

參、啟示：《大數據》與我

世界衛生組織（WHO）渥太華憲章定義的健康係指「身體的、心理的及社會的達到完全安適狀態，而不僅是沒有疾病或身體虛弱而已」，由此可知衛生醫療體系是健全社會的基礎。研讀本書應用巨量資料改進衛生醫療政策的實例，引發筆者對相關研究的興趣。以下謹就「應用巨量資料建立傳染病監測系統」與「應用巨量資料分析改善我國全民健康保險制度」兩大主題，分析我國現有制度及闡述未來發展趨勢之微見如下：

一、應用巨量資料建立傳染病監測系統

疾病監測系統向來被視為「流行病學的智囊泉源」，與公共衛生實務息息相關。利用監測系統即時搜集並分析疾病資訊，散發至各級衛生單位及醫療院所，可作為策劃、執行及評估公共衛生之依據。歷經 1918 年西班牙流感、1957 年 H2N2 流感、1968 年 H3N2 流感、1997 年香港與 2003 年東南亞爆發 H5N1 禽流感、2009 年新型 H1N1 流感、以及 2013 年中國 H7N9 禽流感的挑戰，伴隨 SARS（嚴重急性呼吸道症候群）與 MERS（中東呼吸症候群冠狀病毒感染症）等新興傳染病衝擊，如今我們更明白防疫工作需要以宏觀全球的角度通盤考量，藉由良好的監測系統整合資訊，發揮防微杜漸的功效，才能降低傳染病對人民生命財產的損害。筆者以身為防疫從業人員的觀點，分析書中介紹谷歌流感趨勢與我國現有傳染病監測系統的利弊如下：

（一）谷歌流感趨勢（GFT）的失準

谷歌流感趨勢（GFT）目前共搜集並預測全球 29 個國家流感疫情，以圖 2 為例，圖中顯示處於夏季的北半球國家流感疫情等級為極輕微，而處於冬季的南半球國家則是呈現中等至嚴重等級的疫情。除了 GFT 外，谷歌也應用同樣的原理於 2011 年開發了「谷歌登革熱流行趨勢」，分析新加坡等 10

個熱帶國家登革熱關鍵字在網路上被搜尋的頻率，藉此預測該地區的登革熱疫情。美中不足的是，目前 GFT 及「谷歌登革熱流行趨勢」都不包含我國的資料，因而無法直接比較我國官方數據與谷歌預測數值的差異。

圖 2　谷歌流感趨勢預測 29 個國家流感疫情之地理分布圖（2014.06.21）
（http://www.google.org/flutrends/intl/zh_tw/）

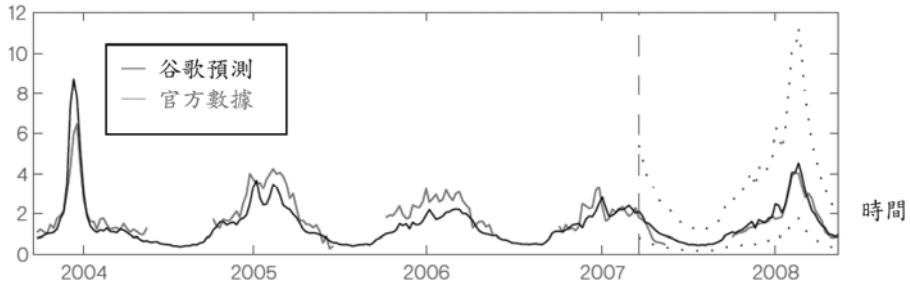


圖 3　2004 年至 2008 年間谷歌流感趨勢與美國官方數據之比較圖

如圖 3 所示，在 GFT 發表的論文裡，谷歌預測類流感盛行率的數值（深色曲線）與美國疾病管制與預防中心公布的數值（淺色曲線）十分相近，兩者的相關性高達 0.9，谷歌比較 2004 年至 2009 年間 GFT 預測值與 18 個國家的官方數據也相當一致。然而當學者持續追蹤 GFT 數值與美國官方數據卻發現，2012 年 11 月至 2013 年 1 月流感大流行期間，

GFT 較官方數據高估了近 2 倍的數值（圖 4）。學者推測可能是因為媒體曾經大幅報導流感，造成網路上搜尋流感關鍵字的人次增加，這些搜尋人次部份來自於對流感議題感興趣的健康人，因而造成誤差。另一方面，由於每年流行的流感病毒株不同，症狀也略有差異。谷歌就曾在 H1N1 大流行時觀察到搜尋關鍵字改變，進而微調 GFT 的模型，以求讓流感預測模型更精準。因此巨量資料分析應該與時俱進，並留意外來因子可能造成的干擾，才能發揮最精確的效用。

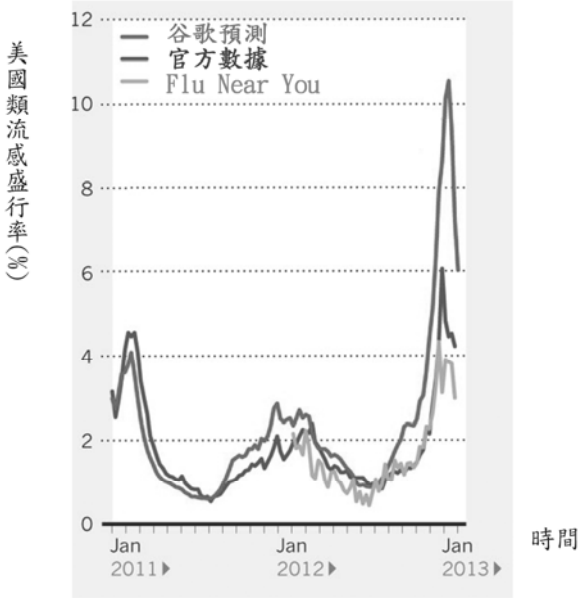
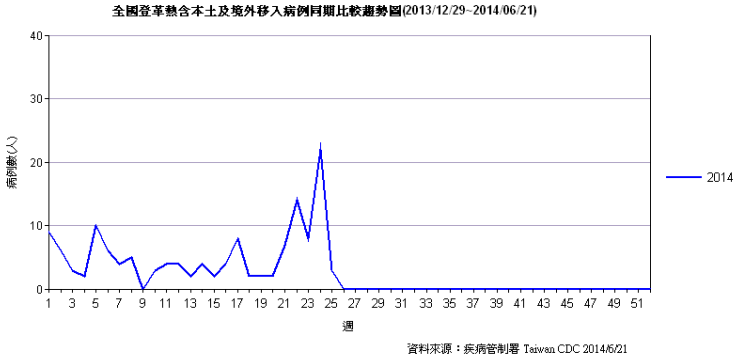


圖 4　谷歌、Flu Near You 與官方數據的類流感盛行率比較圖

（二）我國疾病監測系統的演進

為加強社區傳染病的通報，並評估傳染病對社區造成之衝擊性，我國於 1990 年 1 月設立「定點醫師監視通報系統」，由各地自願加入的定點醫師針對當地類流感、腸病毒、腹瀉等傳染病進行監測，每週通報監測資料。透過分析該通報資料，衛政機關可判斷當時傳染病疫情變化，早期偵測傳染病的發生，並適時提供警訊。隨著我國全民健康保險卡 2003

年全面轉換為 IC 卡，民眾就醫的巨量資料可即時上傳至「全民健康保險研究資料庫」，我國衛生福利部即著手整合「全民健康保險研究資料庫」與具有電子地圖功能的地理資訊系統（GIS）資源，建構出「即時疫情監視及預警系統（RODS）」。藉由全國 170 家責任醫院進行網際網路傳染病通報，每日即時傳送病患的資料至 RODS，以期及早發現社區中可能爆發的傳染病，並建立疾病的流行趨勢及流行預測。衛生福利部疾病管制署每週分析 RODS 資料，進行疫情流行趨勢研判，並於網站上公布疫情統計圖及病例地理分布圖。



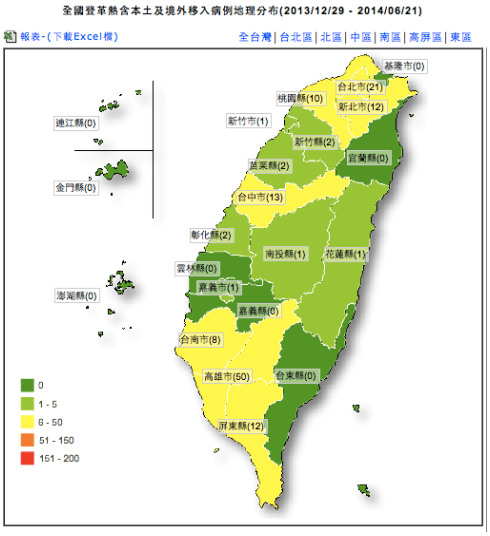


圖 5　疾病管制署網站公布之我國登革熱疫情統計圖（上圖）及病例地理分布圖（下圖）

我國疫情監測系統擁有全民健康保險研究資料庫之優勢，已具備收集資料並進一步將其轉換成有用資訊的能力，然而目前傳染病通報系統缺乏對醫院端主動即時的反饋，醫院端仍需登入系統逐筆查詢個案資料才能得知送驗檢體之結果。此外，由於全國性預防接種資訊管理系統（NIIS）與傳染病通報系統亦未串連，衛生機關也無法即時得知個案的疫苗接種紀錄。有鑒於此，衛生福利部於民國 101 年起推動健康雲計畫，其中防疫雲計畫整合法定傳染病通報系統、疫情倉儲資料系統、實驗室傳染病自動通報系統與跨院所實驗室資料交換平臺等雲端資料，將表單統一簡化，透過電子病歷自動上傳減少人工作業，並提供醫療院所通報訊息的即時反饋，使疫情掌握可更迅速、簡便（圖 6）。除了防疫雲計畫之外，衛生福利部的健康雲計畫也推動醫療雲、保健雲、照顧雲 3 項雲端醫療整合計畫，預計整體計畫將於民國 105 年建置完成，未來可為民眾帶來更優質的健康照護服務。

圖 6　防疫雲執行流程

資料來源：防疫雲計畫辦公室/緯創軟體股份有限公司。

（三）他山之石可以攻錯

防疫視同作戰，醫護人員與各級衛生機關防疫人員莫不競競業業全年無休地監控傳染病疫情。監測系統有如狼煙，如果能夠靈敏即時的傳遞疫情，不但可以提出預警，更有助於後勤防疫物資的整備。將我國疫情監測系統與谷歌流感疫情趨勢的優缺點製成表格比較（表 1），可發現我國疫情監測系統除時效性與民眾參與度略遜於 GFT 外，其餘功能已可與之抗衡，甚至優於 GFT。

表 1　我國傳染病監測系統與谷歌流感疫情趨勢的優缺點比較

特性	我國疫情監測系統	谷歌流感疫情趨勢（GFT）
資料來源	每日收集醫院通報資料	每日收集關鍵字被搜尋次數
時地分析	有	有
特異性	高（確診病例才公布）	較低（易受疫情新聞影響）
病原鑑定	有	無
時效性	快	更快（涵蓋未就醫個案）
民眾參與	較少（醫師才能通報）	多

資料來源：筆者自行整理。

除了 GFT，2011 年波士頓兒童醫院、美國公共衛生協會及薛氏應對全球威脅基金會（Skoll Global Threats Fund）共同創立的 Flu Near You（直譯：流感靠近你）平臺也值得我國借鏡。該平臺每週發送電子郵件給美國與加拿大兩國加入計畫的自願者，調查個人及家庭是否有流感症狀與疫苗接種情形。經由分析 4 萬 6 千個以上自願者提交的資料，Flu Near You 也在網站上公布即時疫情地理分布圖與疫情趨勢曲線圖。相較於 GFT，該平臺 2012 年至 2013 年的監測資料更接近官方公布的數值（圖 4 淺色曲線）。除了蒐集並分析巨量資料，Flu Near You 也同時提供社區民眾衛生資訊，民眾只要在網站上輸入所在地的郵遞區號或住址，就可以獲得該地流感疫苗接種設站的資訊。而我國疾病管制署網站僅提供各

地方衛生局流感疫苗及抗流感藥物配置點清單連結，民眾如果要了解相關資訊，必須還要從各衛生局提供的醫療院所地址清單逐筆搜尋才能找到相關資訊，相較於 Flu Near You 便捷直接的方式，我們還有很大的進步空間。

二、應用巨量資料分析改善我國全民健康保險制度

我國憲法增修條文第 10 條第 5 項明定國家應推行全民健康保險（以下簡稱全民健保），並促進現代和傳統醫藥之發展。爰此，我國於民國 94 年 3 月 1 日開辦全民健保，如今邁入第 19 個年頭，全國 99.9%以上民眾皆已納保，特約醫療院所涵蓋率亦達 93%，大幅改善醫療資源的普及性與就醫可近性。全民健保施行至今，國人的平均餘命由 74.5 歲延長至 79.9 歲，然而隨著人口老化，三高等慢性病盛行率增加，醫療服務費用支出亦水漲船高，因而促成二代健保支付制度的多元改革。

我國參考美國等先進國家的經驗，於民國 103 年起全面推動以住院診斷關連群（Diagnosis Related Groups，簡稱 DRGs）支付除癌症、精神病、罕病等重症外所有住院疾病，並依照病患本身特質與耗用的醫療資源，將住院疾病分為 967 個 DRGs。DRGs 制度以過去醫界提供服務之數據為基礎，計算同組疾病未來給付給醫院的費用，一旦醫院提供超過 DRGs 給付上限總額之服務，健保將僅支付超額部份 80%費用。雖然 DRGs 可以精算醫療成本，避免浮濫醫療造成浪費，卻可能排擠重症與長期住院個案，導致個案產生住院後併發症，甚至增加離院後再急診或是再住院的風險。衛生福利部中央健康保險署分析民國 99 年試辦 155 項 DRGs 群組一年之成效，也的確觀察到住院後併發症、個案離院後再急診率與再住院率 3 項預後指標都上升的現象。

美國是首先推行 DRGs 制度支付醫療服務的國家，本書介紹微軟研究院與醫星華盛頓醫學中心合作案，該中心以微軟開發的 Amelga 軟體分析個案醫療紀錄，希望能找出 DRGs 制度造成再住院

率上升的原因。透過相關性分析，研究人員發現有特定症候群（例如鬱血性心衰竭）的病人再住院率特別高。但出乎意料的是，他們還找出病人心理狀態與再住院率之間的相關性—心情鬱悶的病人再住院機率較高。如果醫院在病人出院後提供心理輔導支持，將有可能降低再住院率及因此衍生的醫療成本。即使歐巴馬總統上任後極力推動美國全民健保，但目前該國健保制度的保險人仍以商業保險公司為主，政府也未主動蒐集分析各業者醫療成本，而是由民間非營利組織成立醫療保健成本研究機構。我國健保資料整合中心與醫療費用審查已全面 e 化，政府可以直接分析全民的就醫資料，若能引進 Amelga 軟體分析 DRGs 制度利弊，對健保制度革新將有如虎添翼的助益。

除了 DRGs 制度，我國二代健保也針對不同形態疾病設計「論人計酬」與「論質計酬」的支付方式。論人計酬鼓勵醫院提供醫療以外的衛教服務，透過宣導預防勝於治療，增進民眾健康。論質計酬則是針對醫療費用大、罹病人數多、照護模式有改善空間之疾病別（例如糖尿病、高血壓、氣喘、肺結核、乳癌等），追蹤評量病患控制疾病的情況，進而提供醫院改善醫療品質誘因。雖然二代健保已逐步改革過去以量計價的支付方式，但世界上沒有完美的制度，我們應善用巨量資料提供的前瞻性思維與選擇，持續檢討並修正，以提供民眾更好的健康服務。

肆、結語：《大數據》的未來

自 2009 年起，美國、歐盟、日本等先進國家紛紛將雲端運算列為重點施政計畫，各國期盼能透過雲端運算提升政府效能並降低成本，協助產業升級轉型與應用服務發展，提升產業競爭力並帶動內需與外銷市場成長。我國亦於民國 99 年研提「雲端運算產業發展方案」，預定設置 15 朵政府雲，加入各國巨量資料競逐的行列。以下謹就我國雲端產業發展的願景與風險管控分述如下：

一、政府雲陸續升空

為了整合各部會資源，簡化行政流程，並提供民眾公開、即時、便利的服務。行政院於民國 101 年 11 月修正「雲端運算產業發展方案」為「雲端運算應用與產業發展方案」，推出讓民眾有感的政府雲（如圖 7）。預計全體計畫將於民國 105 年完成，期許能達成三大願景：

（一）推動民眾有感的政府雲端應用，促成雲端運算觀念與技術普及化，發揮雲端運算建設的綠色節能與成本節約效益。

（二）奠基於資通訊產業，轉型升級為雲端運算產業，讓我國成為具技術自主能力，可提供雲端系統、應用軟體、系統整合與服務營運之技術先進國家。

（三）普及雲端運算應用，發展臺灣成為政府、企業與個人高度使用雲端服務之先進雲端應用典範輸出國。

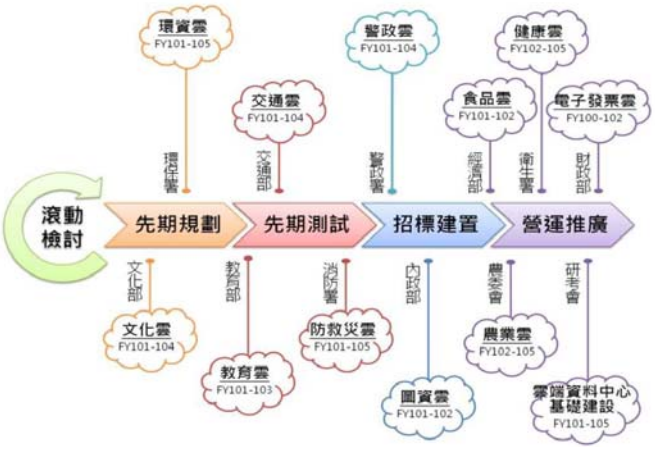


圖 7　政府雲與推動期程

資料來源：雲端運算應用與產業發展方案 103 年 7 月核定本。

「雲端運算應用與產業發展方案」靈活利用政府蒐集的巨量資料，除了提供民眾便捷的服務，並促進產業升級，還能縮短城鄉差

距，使偏鄉共享巨量資料創造的價值。「醫療雲」建構健保雲端藥歷與電子病歷查詢系統，醫師可跨院查詢以往病人用藥紀錄、檢驗結果，降低藥物過敏風險，並減少重複檢驗造成的浪費。偏鄉民眾在城市大醫院做完檢查後，可就近至在地衛生所回診看報告追蹤，無需再舟車勞頓至城市回診，因此彌補了偏鄉醫療設備的不足。「教育雲」也導入偏鄉課輔系統，讓偏鄉學生能無縫接軌享受最新的學習資源。

為了促進我國資通產業的發展，並整合資訊設備的規格，「雲端運算應用與產業發展方案」採用先軟後硬的方式，產官學合作建構雲端平臺與軟體研發，再統一規格採購相應的硬體設備。以「防災雲」為例，我國藉助谷歌完善的全球地圖系統，並整合中央氣象局、行政院農業委員會水土保持局、經濟部水利署、交通部公路總局以及國家災害防救科技中心等重大災害防治單位提供的資訊，成立了「Google 臺灣災害應變資訊平臺」。民眾可透過「臺灣防災地圖」介面從地圖上一目瞭然同時比對氣象雷達回波、颱風路徑、土石流警戒及避難所、公路路況等防災資訊。一旦發生災害，只要在谷歌首頁上蒐尋颱風、地震等關鍵字，或是在智慧型手機上安裝「Google 災害警示」軟體，就可以獲得最新的災情警示資訊。

二、水能載舟，亦能覆舟

「雲端運算應用與產業發展方案」展現了政府運用巨量資料的凌雲雄心，但在共享資訊創造價值之前，必須先保障資訊安全，才能防止民眾的個人隱私甚至國家機密遭不肖駭客竊取。資安事件層出不窮，小至個人電腦資料遭駭客鎖定勒索，大至企業、國家保管的民眾個資被盜取。以下前車之鑑值得我們引以為戒：

（一）美國第二大零售商塔吉特曾經是商界運用巨量資料獲利的標竿，但 2013 年末該公司爆發近 4000 萬筆消費者信用卡資料遭竊的事件，重創其商譽，即使後來亡羊補牢改正系統，塔吉特 2014 年第一季利潤仍較同期減損了 46%。

（二）2014 年 4 月 7 日資安專家發現網路加密軟體函式庫 OpenSSL 有一隻蟄伏超過 2 年以上的網路臭蟲「Heartbleed」（直譯：心在淌血），駭客能利用它竊取用戶帳號、密碼、電腦資料、甚至加密金鑰。「Heartbleed」被認為是近 10 年來最嚴重的網路風暴，不僅 Google、Yahoo、Facebook、Amazon 等大型網站連忙修補漏洞，加拿大國稅局也證實有 900 個民眾網路報稅資料因此遭刪除，美國健保局更史無前例的強制 800 萬以上使用者重設網站密碼。我國行政院資通安全辦公室雖然在 4 月 10 日就對各機關發出資安警告，但卻遲至 4 月 30 日才公布後續漏洞填補的清查結果，顯示我們面對資安事件的處理效率尚待加強。

即使目前我國公開資料集的資訊已經過匿名化處理，承辦人員亦需要通過晶片卡認證才能查詢民眾個資，政府身為資料使用者仍需善盡保障個資安全的責任，留意系統可能有百密一疏的後門程式存在。雖然近年來兩岸互動漸趨和平，但大陸對臺的文攻武嚇從未間斷。試想，如果癱瘓香港電子公投平臺與港臺兩地蘋果日報網站的駭客轉而攻擊我國政府網站，我們有防禦能力與之對抗嗎？

我國過於仰賴委外廠商維護資訊系統，間接造成機關對資安管控約束力不足。雖然規模夠大的機關會建置資訊人員，但現有的人力光是電腦維修與內部資訊系統維護已疲於奔命，實在無暇再分神管理委外系統。筆者認為我們應朝以下 2 個方向努力：

（一）參酌本書「演算學家」的機制，於各機關增設專責資安人員。除了具備資訊管理能力外，資安人員應盡量了解機關執行的業務，作為承辦人員與委外系統之間的橋樑，從使用者導向出發，設計出更完備的資訊系統。

（二）承辦人員應熟悉業務所使用之資訊系統，並以身作則善盡保管資料的責任，不該將資安保護視為專屬委外廠商與資訊人員的工作。使用公務電腦應定期更新補強系統漏洞，並對機

敏資料加密，不點擊釣魚網址或下載來路不明的檔案，以防駭客的社交網路攻擊。

三、勿忘謙卑與人性

已故科幻小說大師艾西莫夫在其經典著作《基地》中自創的「心理史學」學說與《大數據》巨量資料理論有許多相似之處。在艾西莫夫的創作裡，哈理・謝頓教授揉合了歷史學、數學、社會心理學、社會學、氣體動力學及統計學，創立了心理史學。心理史學可根據大規模人類活動數據，推測未來人類群體對特定社會與經濟產生的反應。巨量資料並不是束縛未來人類的可能性，讓我們成為過去行為的囚徒，而是提醒人們未雨綢繆。如同謝頓教授以心理史學預測出銀河帝國即將滅亡後，隨即力挽狂瀾推動「謝頓計畫」，以求將帝國衰滅後人類歷史將持續 3 萬年洪荒時代的宿命縮短為 1 千年。或許將來有一天巨量資料可如艾西莫夫的創作般，預測未來可能發生的災難，指引人們趨吉避凶。

《大數據》作者之一的麥爾荀伯格日前應邀來臺演講的結語提到：「資料終究只是真相的影子，因此，總是不完美的，也總是不完整的。當我們走進這個令人興奮的大數據世界，我們也需要做到謙卑和人性。」巨量資料提供的只是參考的選擇，不是絕對的真理。政府操作國家這個龐然機器，任何政策皆牽動許多群眾的權益，因此更應該秉持謙卑的精神。雖然臺灣是個蕞爾小島，沒有豐富的天然資源，但我國擁有許多優秀資訊人才，實力足以與歐美日等先進國家競爭。且讓我們走出傳統代工的既定模式，循著巨量資料建構的天梯，以虛懷若谷的胸懷站在雲端，引領臺灣走向國際！

參考文獻

一、中文書籍

1. 《微生物學史 開天闢地的醫學拓荒者》（修訂版，2013），第 3 頁—許清曉醫師著，合記圖書出版社

2. 《病毒的故事》（2003），第 100 頁—徐明達著，天下文化
3. 《世紀首疫／H1N1 新型流感大流行工作紀實》（2011），第 19-20 頁—行政院衛生署疾病管制局
4. 《生物統計學》（第六版，2007），第 107 頁—楊惠齡、林明德著，新文京開發出版股份有限公司
5. 《被新聞出賣的世界「相信我，我在說謊」一個媒體操縱者的告白》（2013），第 233 頁—萊恩・霍利得著，張嚴心譯，臉譜出版社
6. 《圖解公共衛生學》（2013），第 12 頁—顧祐瑞著，五南出版社
7. 《公共衛生學》（修訂四版，2008），第 16 章 傳染病流行的偵測與防治原理—陳拱北預防醫學基金會主編，健康文化事業股份有限公司出版
8. 《禽流感大戰疫》（2005），第 14-30 頁—徐明達著，時報文化出版企業股份有限公司
9. 《基地》（1951）—以撒・艾西莫夫著，葉李華譯，城邦文化股份有限公司奇幻基地出版事業部

二、中文期刊

1. 我國「定點醫師監視通報系統」成果回顧，疫情報導第 26 卷第 6 期（2010 年 3 月 23 日），第 81-86 頁—行政院衛生署疾病管制局
2. 101 年傳染病統計暨監視年報（2013 年 12 月出版），第 74-80 頁—衛生福利部疾病管制署
2. 健保臺灣版 DRGs 提升醫療效率成果看得見，全民健康保險雙月刊第 91 期（民國 100 年 5 月號），第 4-6 頁—行政院衛生署中央健保局

三、英文期刊

1.Detecting influenza epidemics using search engine query data, Nature 457, 1012-1014 (19 February 2009)
2. When Google got flu wrong, Nature 494, 155-156 (14 February 2013)

四、媒體報導

1. 風傳媒 http://www.stormmediagroup.com/
「被遺忘權」歐盟判決 Google 須遵守（2014 年 5 月 13 日）
2. 自由時報 http://www.ltn.com.tw/

谷歌受理網路「被遺忘權」首日湧入萬筆（2014 年 5 月 31 日）

3. iThome http://www.ithome.com.tw/

(1) 行政院衛生署打造健康雲 建立人民健康資料總歸戶（2012 年 11 月 2 日）

(2) 教育部電算中心打造教育雲 IaaS 讓教學資源整合再利用（2012 年 11 月 2 日）

(3) Google 結合政府開放資料 打造臺版防災地圖（2013 年 7 月 19 日）

(4) OpenSSL Heartbleed 漏洞危機特別報導（2014 年 4 月）

4. 公共電視 http://www.pts.org.tw/

公視全球現場 深度週報—美資安問題（2014 年 3 月 1 日）

5. 新唐人中文電視台 http://www.ntdtv.com/indexb5.html

大數據時代來臨權威：拒絕數據獨裁（2014 年 6 月 11 日）

五、網頁資料

1. 維基百科中文版 http://zh.wikipedia.org/wiki/

2. 谷歌流感趨勢中文版 http://www.google.org/flutrends/intl/zh_tw/

3. 美國資訊公開平臺 https://www.data.gov/

4. 美國白宮 http://www.whitehouse.gov/

5. 我國政府資料開放平臺 http://data.gov.tw/

6. 谷歌登革熱流行趨勢中文版

http://www.google.org/denguetrends/intl/zh_tw/

7. 衛生福利部疾病管制署—傳染病統計資料查詢系統

http://nidss.cdc.gov.tw/

8. 防疫雲計畫辦公室 http://www.taiwan-nndss.org/

9. Flu Near You https://flunearyou.org/

10. 衛生福利部疾病管制署 http://www.cdc.gov.tw/

11. 衛生福利部中央健康保險署全球資訊網 http://www.nhi.gov.tw/

12. 衛生福利部統計處 http://www.mohw.gov.tw/cht/DOS/

13. 行政院科技會報 http://www.bost.ey.gov.tw/