



# 以大數據挖掘主計資料金礦

未來人類社會出現以資料為驅動力的經濟型態，為因應大數據時代的來臨，本文係探討主計機關應如何運用大數據概念，設法挖掘主計資料之金礦，提升主計政策的品質，以強化行政效率。

丘昌泰（元智大學大數據與數位匯流創新中心、社會暨政策科學學系教授）

## 壹、大數據產業的崛起

近年來，大數據被產業界視為第五波科技革命後的明星產業，未來人類社會將出現以資料為驅動力的經濟型態（data-driven economy）。根據國際數據公司<sup>1</sup>（International Data Corporation, IDC）的調查：大數據科技與服務市場是一個快速成長、產值高達數百億美元的全球新興產業，每年成長率約為 27%，2017 年預估產值將高達 324 億美元，約等於過去資訊通訊科技總產值的六倍之譜。McAfee & Brynjolfsson（2012）指出：大

數據正在掀起管理革命，企業首先必須設法改變傳統決策制訂文化—從“我們想些什麼”（What do we think？）轉變為“我們知道些什麼”（What do we know？），企業必須充分利用大量的客戶資訊，客觀地瞭解大量數據所呈現的意義與價值，然後做出妥適的企業決策，快速地提升企業的營運效益<sup>2</sup>。大數據不僅被應用於私領域，公部門也快速地加以引用，Robertson（2012）指出：以健康產業為例，美國每年大約創造 30 億元產值，大數據已成為「大商機」。事實上，大數據存在於各個領域，包括醫

療健康、食品安全、氣候氣象、能源與環保、智慧交通等，在不違反個資隱私保護前提下，如何妥善利用這些巨量資料，挖掘有價值的決策資訊並形成決策知識，已成為當前政府必須面對的一項「大挑戰<sup>3</sup>」。

本文旨在探討為因應大數據時代的來臨，主計機關應該如何運用大數據概念，以提升主計政策的品質、強化行政效率？本文首先檢視大數據的基本概念，其次探討行政院運用大數據分析的背景與內涵，最後呼籲公務員應該順應此一趨勢，設法挖掘主計資料之金礦。

## 貳、大數據概念的釐清

大數據，或稱海量資料、巨量資料，乃近年來學術界與實務界熱烈討論的新課題。Viktor Mayer - Schönberger and Kenneth Cukier (2013) 於 2013 年出版「大數據<sup>4</sup>」一書，很快地名列亞馬遜書店暢銷書排行榜之一，執教於英國牛津大學，專研網路治理與管制 (Internet Governance and Regulation) 的 Viktor Mayer-Schönberger 及經濟學人 (The Economist) 資深編輯 Kenneth Cukier 指出：今天人類社會從科學到醫療保健、從銀行到網路，涵蓋的行業各式各樣，講的都是同一件事－世界上的資料正在迅速增長，不僅超過電腦機器處理的容量，甚至遠超過我們的想像：Google 每天得處理超過 24PB 的資料 (1 petabyte 等於 1,000 terabyte)，約等於美國國會圖書館所有紙本資料量的數千倍。YouTube 服務的用戶數達八億用戶，每秒上傳的影片長度超過一小時。

臉書每小時上傳一千萬張新照片，按讚或留言者至少超過三十億次，資料量之大難以想像。

何謂大數據？根據美國白宮總統辦公室 (Executive Office of the President, the White House) 公布的大數據報告<sup>5</sup> (The Big Data Report)：大數據的概念需視你是何種專家而定，多數的定義都顯示：由於當前雲端科技的發達足以快速地擷取、累積、利用與處理個人資料，以致於出現數量大、速度快與種類雜的資料型態，無論是政府與企業都應該加以利用，以提升決策品質。基此，甚至有學者說：大數據是二十一世紀人類所發現的「新石油」。

大數據與傳統的數據是不相同的，有人以 3V<sup>6</sup> 來表示其應具的特徵 (McAfee & Erik, 2012)：

### 一、資料量相當龐大 (Volume)

透過各種蒐集資料的新科技工具，如感應器、數位相

機、GPS 與其他觀察人類行為的資通訊科技，使得當前資訊社會蒐集、儲存與處理資料的成本愈來愈低，因此我們正處在一個資料幾乎無時無刻被蒐集的世界。據統計，2012 年世界每天產生的資料量約為 2.5 exabyte (EB)，這個數字大約每 40 個月左右就會增加一倍。

### 二、資料產生速度極快 (Velocity)

網際網路每秒流傳的資料量，比二十年前網際網路儲存的資料快了太多。目前人類藉著網路驅動的工具、可攜式行動裝置到現代化的感應器，可以完整的追蹤人們所遺留的「數據足跡」，而且速度甚快；例如，當人們上線瀏覽網頁或按讚已被完整的記錄下來，網站經營者能夠準確地追蹤每位使用者的線上活動與行為；或當人們以行動裝置使用 GPS 時，其位置已被確實地掌握住；甚至人們在社群網站上的留言，都被完整地記錄下來。



### 三、資料類型相當多元 (Variety)

根據美國總統科技顧問委員會 (the President's Council of Advisors of Science & Technology, 2014: 4) 的報告，大數據資料可以分為兩大類：一是原生數位資料 (born digital data)，這是指透過電子郵件、網路瀏覽、GPS 定位系統或其他電腦資料處理系統，直接從有形世界中轉化為數位化的資料。二是感應資料 (sensors' data)，這是透過錄影機、手機、照相機或無線感應器等所接收的資料。若能針對這兩類數據進行數據融合 (data fusion)，產生有價值的資訊，將對人類社會產生莫大的影響。從資料的次序性而言，大數據可以分為結構化與非結構化資料 (structured and unstructured data)，前者係指有規則可循，可以用統計或數學演算方式加以處理的大量資料，後者則指毫無規則可循的、隨機的資料類型，如社群媒體

資料。基此，如何整合該兩大資料類型，並加以分析，以得到有用的資訊將成爲人類社會最大的挑戰。

### 參、行政院推動大數據分析的政策背景與內涵

由於大數據研究爲相當新穎的課題，故檢視臺灣有關大數據的研究文獻相當之少，大多數文獻都是在呼籲學者應加強對大數據的應用，如賴鼎陞<sup>7</sup>以大數據分析博物館觀眾行爲，張文貴<sup>8</sup> (2013) 認爲品質管理應該積極面對大數據時代的變革趨勢，呂宗學與蘇慧貞<sup>9</sup> 探討海量資料分析在醫療照護領域的應用；蔡博坤<sup>10</sup> 探討智慧聯網時代巨量資料法制議題研析—以美國隱私權保護爲核心。在大數據分析方法方面，如 Jordan, John M. 與林共進<sup>11</sup> 呼籲應該加強統計分析能力等。

毛治國擔任行政院長，在張副院長善政的推動下，要求全體閣員「換腦袋」，加強

與網路族群的溝通，並結合民間力量推動嶄新的網路平臺，協助解決政府與網民溝通的困境。因此，毛揆射出「科技三箭」—開放資料、大數據與群眾外包，希望各政府部門能夠妥善運用網路新科技，運用政府資料，以創造有感施政。其次，行政院從科技會報、國發會與資策會調派三組人馬，進駐各部會「挖寶」兼「健診」，毛式科技三箭務必盡快能夠開花結果。

毛院長的大數據治理概念是以資料開放 (Open Data) 爲中心，採取大數據與群眾外包方式，將政府所開放的資料進行分析，以產生有價值、創新的決策資訊，俾供政府制定讓人民有感的公共政策。基此，行政院乃將 2015 年定調爲「開放資料深化應用元年<sup>12</sup>」，釋出政府資料供民間使用，並且仿效英美等作法—以開放爲原則，不開放爲例外；不收費爲原則，收費爲例外。開放資料格式包括：PDF、JPG、DOC、Excel、CSV、XML

等，並邀請民間社群，如 Open Data 聯盟協助轉換政府資料格式，讓缺乏人力、經費的部會加速釋出資料。截至 2015 年 11 月總共開放 12,969 個資料庫，中央政府就釋放了 12,805 個，佔 98%，偏重於中央政府機關，其中尤以法務部開放資料達 2,683 筆資料最多，約占五分之一<sup>12</sup>。

張副院長善政出席全球大數據專家麥爾荀伯格（Viktor Mayer-Schonberger）訪臺論壇時，針對政府近 1 年多來，力推 Open Data、Big Data 和「工業 4.0」、「生產力 4.0」政策時指出：傳統產業企業主，可能認為大數據和他沒關係，也不會特別去觀察大數據對產業趨勢的影響，結果在不知不覺中，就漏搭了這班推動轉型列車。行政院張副院長善政乃呼籲，中小企業主必須要「換腦袋」，千萬不要錯過產業轉型的機會<sup>13</sup>。作為資料提供者的政府部門而言，自然應該遵循行政院的指示，努力推動大數據分析的應用。

## 肆、主計機關因應大數據時代的可行策略

根據政府資料開放平台<sup>14</sup>的統計，臺灣開放資料集總數為 12,702 筆，法務部名列第一名，開放了 3,191 筆，主計總處共發布 485 筆，居第八名。在瀏覽量方面，交通部居第一名，次數達 572,317 次；主計總處瀏覽次數為 204,147 次。至於下載量則以文化部為第一名，下載量為 186,491 次，主計總處為 26,585 次。為因應行政院推動資料開放平台之趨勢，以供學術界與實務界進行大數據分析，謹依據前述有關大數據概念，提出下列政策建議作為決策參考：

### 一、開放資料的法源依據不是《政府資訊公開法》，而是回應全球重視開放政府、透明政府的潮流，開放的是具有大數據分析價值的礦脈資料

很多承辦人係根據民國 94 年 12 月 28 日公布的《政府資訊公開法》第四條：「本法所稱政府資訊，指政府機關於職權範圍內作成或取得而存在於文書、圖畫、照片、磁碟、磁帶、光碟片、微縮片、積體電路晶片等媒介物及其他得以讀、看、聽或以技術、輔助方法理解之任何紀錄內之訊息。」因而《政府資料開放平台》開放許多不具分析價值的政府資料，例如組織制度、職掌人力、各種業務報告、通訊住址、電話總機等不是大數據分析所需要的資料型態。以英國政府為例，公布內容包括各機關首長的薪資、每月政府經費支出概況等。

### 二、《中華民國統計資訊網》只是一般的「數據平台」，而非「大數據平台」

或謂《中華民國統計資訊網》長期以來，已經建立了國人皆知的數據平台，且使用量甚多，不知有何必要再開放大



數據資料庫？然而，這些資料庫只是傳統的數據分享平台，數據單元是年度、縣市、機關等，數量不夠龐大。根據大數據分析的定義，關鍵就在於「大」這個字，係指數量龐大到幾乎與母體一樣，當數據大到母體一樣，則進行抽樣調查就沒有意義；若能從如此龐大的數據庫中找到決策資訊，則對於行政機關制訂公共政策將有極大的助益。截至目前為止，臺灣公布最完整的資料庫是全民健康保險資料庫<sup>15</sup>，由於全民健保納保率已達 99% 以上，健保資料庫幾乎等於臺灣社會的母體，成為醫藥衛生相關領域中最具有代表性的實證資料。

### 三、提高數據資料的精細度，至少要到「奈米級」的原始資料

不少行政機關提供的資料密度不高，以至於該資料沒有任何再研發的價值。《中華民國統計資訊網》的數據單位大

多是「年度」、「機關」、「縣市」，這些單位仍嫌太粗略，建議應視採集統計資料的性質，提供更為細緻的單位資料，例如，提升到以「鄉鎮區」為單位的資料等級，若能提升到的「鄰里」或「個人（全民健保資料庫）」，則已屬「奈米級」資料格式，必然更能挖掘出資料的黃金。就資料開放格式而言，盡量以 excel 為原則，如此欲轉成統計分析軟體較為方便，惟不少機關用 word 或 PDF 格式就相對麻煩許多，甚至不少資料必須重新 key in，這對於分析者而言，乃是相當困擾的事情。

### 四、主計資料庫必須與其他部門之資料庫交互運用與比對，以提升其分析的附加價值

若「單獨」分析主計資料庫，其用處未必很大，這是何以任何機關決策時都會用到統計資料，但主計資料庫之使用率卻不高？大數據的特徵之一

為「資料多元性」，透過大數據分析工具予以整合，如此將更全面地瞭解數據所呈現的意義。例如，與預算有關的單位是立法院及其幕僚機構預算中心，立法委員審查預算的意見都可在立法院議事紀錄中找得到，其資料型式是屬於非結構類型，必須經過運用資訊耙梳程式予以整理，並與主計總處的預算科目與數字相勾稽，如此則對於未來審查預算將有極大助益。與決算有關的機關則是審計部，審計部對於行政院의年度審計報告都會提出許多寶貴意見，若能與預算相勾稽，等於從政策規劃、預算籌編、預算執行與決算審計都有所聯結，助益必然甚大。

其他可以交叉運用或比對的資料庫甚多，均可請分析單位發揮其創意，俾從相互比對中發現特殊的社會現象，進而制訂政策予以解決。例如，人文社會經濟地理統計資料與罹患癌症者分布之關聯性，雖需要更多的臨床研究找出原因，

但也指出若干國民健康政策方向，值得深入追蹤與調查。

## 五、主計資料若要挖掘出黃金，建議應以研究計畫方式委請專家學者進行探究

大數據分析是一項耗費精力、時間與金錢的工作，依照目前「群眾外包」作法，只是將資料免費開放給民眾使用，然後衷心盼望民眾能夠自動自發的加以分析，並得到創意成果，這無疑是緣木求魚。君不見政府開放資料平台中，所謂活化應用的成果大都屬於相當初階的粗淺成果，將資料鍵入後，運用搜尋工具找出位置，如停車場資料等。為了加強主計資料庫的深度，建議與學者專家、學校或研究機構以研究計畫方式簽約進行系統性的探究，如此可以得到比較深入的成果，另一方面則可確保資料的運用符合個人資料保護法的原則。

## 伍、結語

很多公務員的疑問是：為什麼要運用大數據分析？政府機關從龐大的數據庫中，或許可以藉著學者專家，提煉出重要的研究發現，而這些發現可能是運用平時作業程序所無法找出的問題盲點，有助於提升政府財政的效能，改善民眾對於政府的不良觀感。例如，財政部曾以大數據調查過去三年全臺 484 萬受薪總人數，依據近三年課稅資料計算的全年薪資所得，100 至 103 年月平均薪資由 47,517 元成長至 48,715 元，3 年來約成長 2.5%。103 年各業別薪資成長以藝術、娛樂和休閒服務業，以及金融保險業成長幅度最高，調幅分別為 9.11% 與 7.71%<sup>16</sup>。這些數據對於政府制訂公共政策是助益的，如此就能制訂出以證據為基礎的政策分析（evidence-based policy analysis），有助於說服社會大眾，並與立法委員進行理性的政策溝通。

然而，我們也必須提醒各位，大數據分析不一定能夠挖掘到黃金，原因甚多，可能是研究者能力不足、研究方法錯誤，但也可能是政府資料本身太過粗糙，並無多大的分析價值。因此，要推動一個成功的大數據分析，需要的配套條件不少，上面五點可供大家參考，期盼運用群眾智慧，能夠順利挖出主計資料的金礦。

## 註釋

1. 國際數據公司（IDC）是全球著名的資訊技術、電信行業和消費科技諮詢、顧問和活動服務專業提供商。IDC 在全球擁有超過 1,000 名分析師，對 110 多個國家的技術發展趨勢和業務行銷機會進行深入分析。請參閱：IDC, Big Data/Analytics 專區，<http://www.idc.com/home.jsp>。上網日期：2015/9/28。
2. McAfee, Andrew & Brynjolfsson, Erik (2012). "Big data: The management revolution," Harvard Business Review, October: 61-69.
3. Robertson, J. (2012, May 17). "The health-care

## 論述 » 專論 · 評述



- industry turns to big data." Bloomberg Businessweek, <http://www.businessweek.com/articles/2012-05-17/the-health-care-industry-turns-to-big-data>.
4. Viktor Mayer-Schonberger and Kenneth Cukier (2013). Big Data: A Revolution That Will Transform How We Live, Work, and Think. UK: Hodder & Stoughton Limited. 林俊宏譯, 《大數據》, 臺北: 遠見天下文化出版股份有限公司。
5. Executive Office of the President, the White House (2014), Big Data: Seizing Opportunities, Preserving Values, May: 2-3.
6. 也有學者認為應該要再加上一個 Value, 而形成大數據的第四項特徵, 因而形成 4Vs, 意指所有大數據分析的最終目的都是期盼從浩如煙海的龐大資料庫中, 設法尋找出有價值的、可做為決策依據的決策資訊。
7. 賴鼎陞: 《博物館觀眾研究新契機—大數據》, 《故宮文物月刊》2014年, 第374期, 第94至102頁。
8. 張文貴: 《積極面對「大數據」時代的變革趨勢》, 《品質月刊》2013年, 第49卷第11期, 第7至11頁。
9. 呂宗學與蘇慧貞: 《海量資料分析在醫療照護領域的應用》, 《臺灣醫學》2013年, 第17卷第6期, 第652至661頁。
10. 蔡博坤: 《智慧聯網時代巨量資料法制議題研析—以美國隱私權保護為核心》, 《科技法律透析》2013年, 第25卷第10期, 第46-62頁。
11. John M. Jordan與林共進: 《Statistics for Big Data: Are Statisticians Ready for Big Data?》, 《中國統計學報》2014年, 第52卷第1期, 第133-149頁。
12. 有關政府開放資料之政策淵源甚早, 張善政擔任政務委員於2012年行政院科技會報辦公室陳報「政府資料開放推動策略」報告就已定案, 請研考會與法規會共同研議。另請經濟部研擬產業發展之各項需求及因應策略, 各部會亦應自民衆的應用面發想, 思考使用端之需求, 而在規劃時也要考慮到機器讀取介面的必要性。院會決議與院長指示, 行政院2012年11月8日第3322次院會議決議, 請參閱: 政府資料開放平台, 2012年1月18日行政院科技會報辦公室舉辦「我國公開資料加值 (Open Data) 推動策略會議」; 2012年10月15日行政院國家資訊通信發展推動小組第31次會議, <http://data.gov.tw/about>. 瀏覽日期: 2015/4/12及2015/11/15。
13. 〈張善政: 中小企業要換腦袋〉, 《中時電子報》, 請上網: <https://tw.news.yahoo.com/%E5%BC%B5%E5%96%84%E6%94%BF-%E4%B8%AD%E5%B0%8F%E4%BC%81%E6%A5%AD%E8%A6%81%E6%8F%9B%E8%85%A6%E8%A2%8B-215006130.html>. 2015/9/28。
14. 根據政府資料開放平台的統計, 請上網: <http://data.gov.tw/>, 本項資料之引用日期為: 2015/9/27。
15. 全民健康保險研究資料庫, [http://nhird.nhri.org.tw/talk\\_05.htm](http://nhird.nhri.org.tw/talk_05.htm), 瀏覽日期: 2015/9/28。
16. 林安妮報導, 〈大數據看加薪金融、藝術好薪情〉, 《經濟日報》, <http://udn.com/news/story/6/1012898>. 2015/9/28。❖