

● 公務機關巨量資料分析應用推動簡介

國家發展委員會資訊管理處 劉宗熹

壹、前言

103年底毛前院長上任時提出「網路溝通與優化施政三支箭」之科技政策，希望透過「開放資料」、「群眾外包」及「大數據」概念促進施政品質與效能，協助政府解決施政重要課題；張前院長善政（時任行政院副院長）接受專訪時亦提及：透過大數據分析，洞悉以前沒看到的現實，進而做出對症下藥的政策。從兩位院長的談話內容明確顯示政府巨量資料分析為公務機關推動資料應用的發展重點。

為因應網路時代新趨勢，藉由網路科技及創意，讓政府機關施政更具前瞻性，行政院於104年7月通過創意臺灣ide@ Taiwan 2020政策白皮書，其任務之一是推動善用巨量資料分析提升政府施政品質，規劃以資料科學研究方法分析跨域資料；另成立行政院大數據技術指導小組，責成經濟部、科技部及國家發展委員會協助政府機關辦理巨量資料分析作業，針對民眾關心議題或重大施政課題，如薪資、學用落差、人力資源運用等議題研析問題成因，俾以提出解決對策。截至105年4月底，大數據技術指導小組架構下，已完成7項巨量資料分析議題，並獲得初步分析成果。有鑑於大數據技術指導小組屬臨時性任務編組，為持續推動政府巨量資料分析作業，行政院資訊通信發展指導小組第37次會議決議增設資料應用推動指導小組，賡續推動公務機關、產業及學研單位發展資料應用相關事宜，大數據技術指導小組解除任務。

貳、推動策略

為協助公務機關培養運用巨量資料分析能力，政府推動巨量資料分析應用係以公務機關與學術單位合作模式，由公務機關從社會安全、經濟發展與環境永續等三面向提出重要施政課題，如傳染病預測、毒藥品防制、物價監測、提升薪資、環境品質預測、畢業生就業情形、新住民二代潛力發展等，由國內學術單位或跨機關成立合作小組方式共同進行資料分析，藉此強化政府施政決策品質，同時培育國內急需之資料科學人才；各項分析議題完成研究後，該項分析作業納入機關常態性業務，作為輔助機關施政之一環，以確保政務推動符合民眾與社會需求，並提升政務推動之政策、危機處理與前瞻創新品質與效能。

一、目前政府推動巨量資料分析作業，以政學合作及機關試辦二種模式進行：

(一) 政學合作：由行政院指定應用領域，由科技部向學界徵集研究計畫，藉

此強化政府施政決策品質，培育國內資料科學人才。

(二) 機關試辦：由行政院指定數項重要施政課題，擇定主、協辦機關成立跨域合作小組，透過試辦方式讓公務機關實際進行辦理資料應用工作，培養相關能力。

為協助各部會辦理推動巨量資料分析領域事宜，行政院於104年5月2日成立「大數據技術指導小組」（如圖1），由副院長擔任召集人，國家發展委員會、經濟部與科技部擔任小組成員，分別就法規環境協調、產業技術輔導及學研領域輔導等面向協助推動政府巨量資料分析，同時確立指導小組架構下各項分析議題進程序（如圖2）；各項議題辦理時限以6個月為期，由主責部會邀請相關機關組成工作小組後，進行各項分析工作。

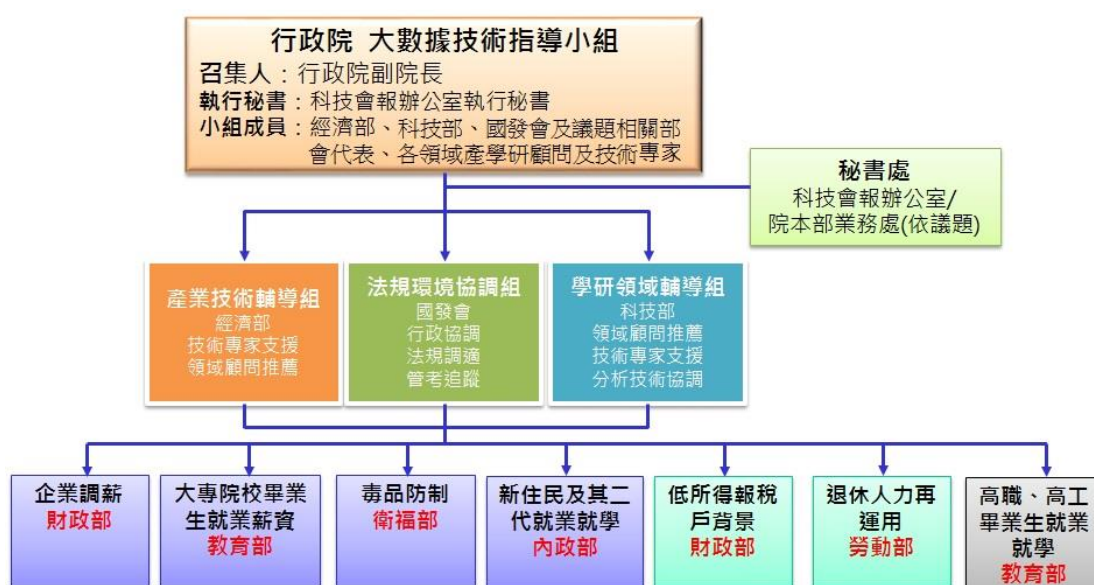


圖 1：大數據技術指導小組架構圖



圖 2：技術指導小組進程序

大數據技術指導小組運作初期指定分析重點以企業調薪、學生就業、毒藥品防制、新住民潛力願景等 4 項議題，隨著上開議題陸續產生初步分析成果，行政院陸續指定「低所得背景分析」、「促進銀髮人力運用及資源整合」與「高工職畢業生生涯發展趨勢」等議題；至此，巨量資料分析應用議題共有 7 項議題。

前揭應用議題之主、協辦機關及與應用議題內容（如表1）：

表1：巨量資料分析應用議題簡介

名稱	主、協辦機關	議題簡介
企業調薪	主辦機關：財政部 協辦機關：勞動部	運用財政部國人薪資所得資料與勞動部勞保投保資料勾稽，客觀反映國人實際薪資。
大專院校畢業生就業薪資	主辦機關：教育部 協辦機關：財政部、勞動部、內政部、國防部、銓敘部、	運用教育部大專畢業生資料、勞動部投保資料及財政部薪資所得資料等分析大專畢業生就業及薪資水準，作為政府規劃教育政策參考。
毒品防制	主辦機關：衛生福利部 協辦機關：法務部、教育部、內政部、勞動部、司法院	彙整分析各毒品防制業務主管機關之業務資料，瞭解毒品使用的輪廓、檢討評估目前的處遇模式，解決當前毒藥品濫用問題，並作為毒品防制政策規劃之參考。
新住民及其二代就業就學	主辦機關：內政部 協辦機關：教育部、勞動部、經濟部、財政部、衛生福利部	分析新住民及其二代在就業、就學與家戶經濟相關資料，透過資料分析洞察不同類型、不同因子的消長趨勢，作為未來切合新住民所需之政策參考。
低所得報稅戶背景	主辦機關：財政部 協辦機關：內政部、勞動部、衛生福利部、原民會、退輔會、農委會、金管會	分析「綜合所得稅所得總額20分位統計分析表」最低分位報稅戶之所得、財產及投資情形，推估其真實收入情況，作為政府施政參考。
退休人力再運用	主辦機關：勞動部 協辦機關：退輔會、教育部、銓敘部、人事行政總處	運用軍、公、教、勞保各類退休人力相關數據資料，找出退休人力可再投入就業市場之潛在勞動力，作為促進退休人力再就業相關政策措施之規劃參考。
高職、高工畢業生就業就學	主辦機關：教育部 協辦機關：勞動部、內政部、財政部、國防部、銓敘部	延續大專院校畢業生就業薪資議題之分析，增列高職畢業生就業狀況與就業後回流就學情形，藉以檢視我國技職體系發展情形。

參、政府巨量資料分析作業要項

大數據技術指導小組架構下推動之 7 項巨量資料分析議題，已於105年4月全數執行完成，並獲得初步分析成果；在各項議題之主、協辦機關與大數據技術指導小組共同合作下，逐漸累積辦理巨量資料分析經驗，本會從中萃取重點工作項目說明如下，提供公務機關參採：

一、政府巨量資料分析三要素（如圖4）

- (一) 資料：依據分析作業所需取得之資料，包括公務機關依職掌而蒐集之資料、政府開放資料、公開資訊或公務機關外部可取得之結構化、非結構化資料。
- (二) 資料運算資源：泛指各類為進行巨量資料分析而準備資料處理、儲存、運算之資訊軟、硬體。
- (三) 資訊科學家：係指具專業領域知識、熟悉統計方法及電腦科學之資料專家；由於公務機關同時具備三項專長之專家非常少，爰政府巨量資料分析作業以團體合作方式，由業務單位、統計單位及資訊單位共同成立工作小組辦理相關工作。

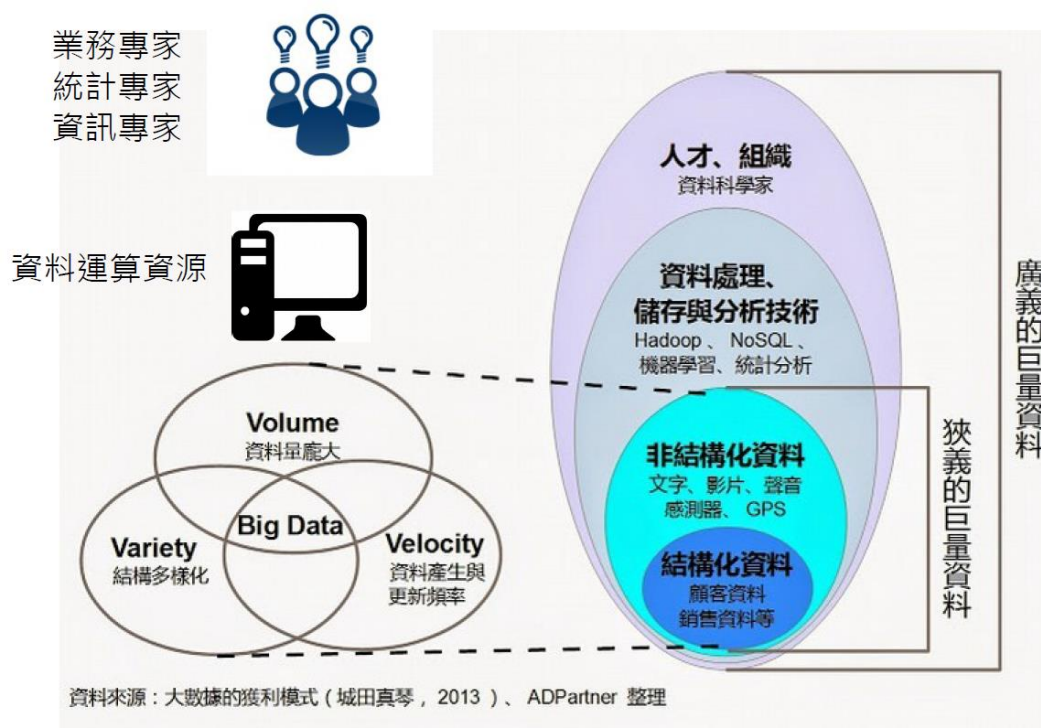


圖4：政府巨量資料分析三要素

二、政府巨量資料分析執行情序（如圖5）

(一) 問題定義

1、釐清政策推動瓶頸或民意輿論焦點

從政策推動、政府業務執行過程遭遇之困境，或蒐集民眾意見、媒體輿論，釐清巨量資料分析執行方向。

2、訂定待解決課題或預定創造之價值

訂定巨量資料分析工作預期達成之目標，進一步研定預計探索之問題，期望找出輔助公務機關解決施政課題之參考。

(二) 前置整備

1、確定分析範圍、資料需求以及分析方法

經過前開問題定義階段已明確得知巨量資料分析作業之目標與預期價值後，依其規劃分析範圍，例如哪些業務領域需要被探索、哪些機關的業務資料可以挖掘出潛在知識、哪些資料串接之後可以找出更多有利分析的資料集等。最後再依據資料性質、問題性質找出可能可以獲得預期結果決定數據分析（統計分析）方法。

2、組成工作小組

邀請相關業務機關（或資料提供機關）擔任協辦單位，除建立業務資料流通與再利用之程序與管理機制外，另協助檢視分析過程與結果之合理性，確保分析方向正確；主辦單位內部則由業務單位、統計單位與資訊單位共組資料科學專家團隊，分別就領域知識、統計方法與資訊技術共同執行資料分析與解讀作業。

3、運算資源與資料資源整備

準備巨量資料分析作業所需電腦運算資源，泛指各項資料科學專家團隊進行資料分析所需之電腦軟、硬體；有關資料資源整備作業，除了蒐集分析所需資料項目外，重點工作係檢視資料內容是否能滿足分析所需，包括資料內容之定義、資料品質、使用限制（如法規規定、時間區間、地區範圍等）。

(三) 執行分析作業

1、資料分類、分級

將資料以各種屬性進行資料分類、分級工作，俾利進行資料分析作業，如地區、年齡、國籍、職業等。

2、資料清理、儲存與運算

去除無效資料（dirty data）後執行資料勾稽、統計相關之資料運算與資料儲存工作，去除無效資料之目的在將資料內容不具意義、錯誤或極端值之紀錄清除，避免造成分析結果異常。

3、跨域資料勾稽與串連正確性

不同領域之資料彼此串聯與勾稽時，必須確定資料之間具有相同資料定義，或資料具有串連可彰顯其正確意涵；例如食品檢驗資料串連原物料檢驗資料，兩者均為檢驗資料，彼此卻無關係。

4、確保資料分析過程之安全與隱私

巨量資料分析經常需要使用各領域之原始資料進行分析作業，然而原始資料如具有機敏性質，資料分析過程必須隨時檢查資料是否儲存不當造成盜用或冒用，而侵害資料當事人權益。

5、解讀資料勾稽結果合理性

解讀勾稽過程產生之數據是否偏離預期，如偏離預期且無法找到合理解釋，必須回頭檢視資料內容是否存在無效資料，導致資料串接結果偏離合理範圍。

(四) 研析產出結果

1、研析分析意義

各項資料勾稽作業完成，並統計出數據發展趨勢或群聚分布後，從業務領域觀點洞悉資料整體呈現意涵，並以落實於業務推動層面，研析分析結果對業務推動之影響。

2、確認結果價值

檢視分析結果是否已達成前所預定達成之目標，或是打破現有業務現況，挖掘出創新價值，進而精進業務發展方向。



圖5：巨量資料分析作業程序

肆、政府巨量資料分析之隱私保護作法

因應國內推動資料應用（如巨量資料分析及開放資料等）而引發民眾對資料隱私侵犯疑慮，行政院自104年3月起著手研議建立個人資料去識別化驗證標準；在蔡前政務委員玉玲督導下，由法務部於vTaiwan平臺與民間社群討論「個人資料去識別化議題」，凝聚民間與公務機關之推動共識；隨後責成經濟部建立個

個人資料去識別化驗證標準（如圖6）。



圖6：個人資料去識別化驗證標準規範文件架構

為確定個人資料去識別化驗證標準確實可行，行政院指示財政部、內政部與衛生福利部，分別就職掌所蒐集之民眾個人資料試辦導入個人資料去識別化驗證標準，其中財政部之「102年度所得稅核定資料」已於104年8月底導入完成，並經過第三方機構驗證後確認導入成功，其他兩個部會刻正執行試辦導入中。未來公務機關必須視法規規定或資料應用內容，適時導入去識別化驗證標準，導入原則如下：

- 一、法規明文規定者（如個人資料保護法第6條）
- 二、擬開放之政府資料含個人資料者
- 三、跨機關流通之資料含有個人資料者

另外責成法務部協助釐清公私部門導入資料去識別化之法律責任，協調司法機關進行個人資料去識別化之法律意見。依據法務部「公務機關利用去識別化資料之合理風險控制及法律責任」之研析報告指出：公務機關保有之個人資料，運用各種技術予以去識別化，而依其呈現方式已無從直接或間接識別該特定個人者，即非屬個人資料，自非個資法之適用範圍。

伍、後續推動方向

因應國際間政府開放資料、巨量資料及數位服務個人化（My Data）等資料應用發展趨勢，藉以提升政府施政效能，並發展資料經濟產業，有必要建立一個溝通、協調並建立合作關係之管道，整合政府端、產業端以及學研端資料應用量能，共同發展我國資料應用相關措施。鑑於大數據技術指導小組為臨時性任務編組，為常態推動公務機關辦理巨量資料分析作業，行政院「國家資訊通信發展推動（NICI）小組」第37次會議決議增設「資料應用推動指導小組」（如

圖7) 接續推動政府巨量資料，該指導小組成立後，大數據技術指導小組解除任務。

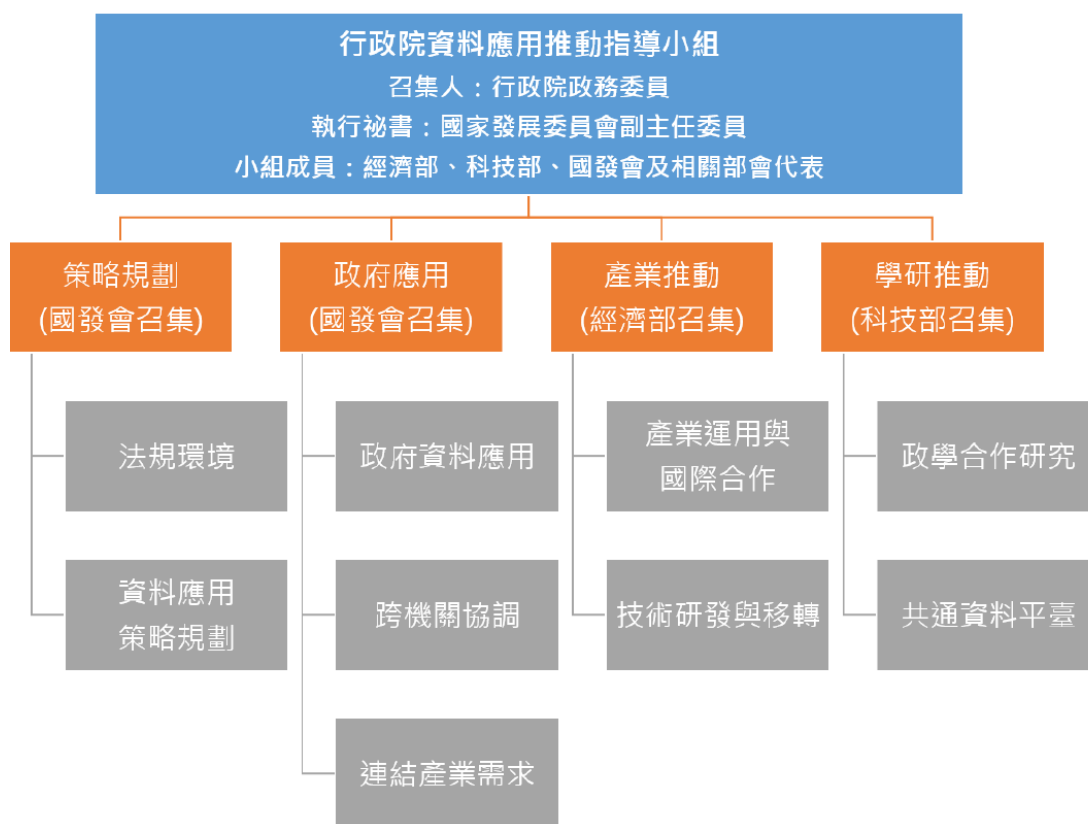


圖7：資料應用推動指導小組組織架構

透過「資料應用推動指導小組」之跨機關、跨業務協調管道，打造政府、產業、學界之合作模式，提升我國資料應用價值，解決政府施政課題，運用個人資料去識別化標準，避免資料應用衍生隱私侵害之疑慮；另外，從法規、人力、技術與資訊資源等面向帶動資料經濟產業發展；最後，建構共用性資料運算平臺合作模式，鼓勵公務機關釋出資料提供學界研究，提升資料應用鏈結綜效。

隨著大數據指導小組架構下推動之7項應用議題結束，各主辦機關已累積相當多推動經驗，後續本會將陸續邀請各議題主辦機關分享議題分析過程與結果，並從實務經驗撰寫分析作業程序，提供公務機關參考及運用。

陸、資料參考來源：

- 一、 創意臺灣 ide@ Taiwan 2020 政策白皮書。
- 二、 行政院「國家資訊通信發展推動（NICI）小組」網站
<<http://www.nici.gov.tw/cp.aspx?n=A9ABC4E4B8529CD7>>
- 三、 財政部大數據薪資分析簡報。

- 四、 教育部 99-101 學年度大專校院畢業生就業薪資分析結果簡報。
- 五、 衛生福利部運用巨量資料分析規劃毒品防制政策-初步成果。
- 六、 內政部 105 年新住民及其家戶巨量資料專案分析簡報。
- 七、 財政部財稅資料低所得報稅戶背景分析。
- 八、 勞動部退休人力再運用分析。
- 九、 教育部運用巨量資料分析高工職畢業生就業及就學情形分析結果。