

【淺談巨量資料】 與生物多樣性監測

Big data and biodiversity monitoring

林大利 / DOLLY / 行政院農業委員會特有生物研究保育中心助理研究員
/ thrush1250@gmail.com

生物多樣性的變化趨勢，需透過大範圍的長期監測才得以參透。圖為布袋度冬的反嘴行鳥群



前言

資料(data)是記錄事實及現象的訊息載體,以數字、文字、圖像、數位等方式呈現(Checkland and Holwell 1998)。電腦的普及與網際網路的發展,開啟了資訊革命,對人類的生活與文化產生極大的改變(Veneris 1990)。30多年前,電腦及網路科學突飛猛進的發展,使各種現象能快速數位化及儲存,讓數位資料大幅成長。資料爆量成長之後,人們對於資料的思維有了轉變,從中挖掘出以往不容易參透的訊息。

大數據(big data),又稱巨量資料或海量資料¹,是指資料量非常龐大,無法在合理的人力及時間內蒐集、儲存、管理、分析及展現,資料量要龐大到一定程度,其功能才會發揮(White 2012)。2007年,全球儲存超過300EB的資料²,其中僅有7%為紙張、書籍和相片等類比資料(Hilbert and Lopez 2011)。依據國際數據資訊公司(International Data Corporation)的估計,2011年全世界儲存至少1.8ZB的資料,2012年則成長到2.8ZB,預估至2020年將達到40ZB。由此可見,數位化的資料量不僅前所未見,成長的速度更是一日千里。

第一次翻開「大數據」這本書是隨意逛書店時巧遇,我對資料科學並不熟悉,隨手翻了幾頁之後便放回原位。不久之後,在國家文官學院的「每月一書」再次遇見「大數據」,便開始悉心閱讀。出乎意料地,「大數據」不僅點燃我對資料科學的興趣,也在業務

及研究上找到新的發展方向。

整體而言,「大數據」讓我初步認識資料科學,若要能夠有效的管理及應用巨量資料,還須具備資料管理、資料庫、程式語言及電腦科學等相關知識與技術。為了進一步學習相關知識,開始悉心鑽研巨量資料,探討挖掘資訊、過濾雜訊及校正偏差的方法,並將成果應用於保育議題,檢視我國生物多樣性的狀態與動態。

專書梗概：由大數據掀起的資訊革命

「大數據」一書由牛津大學網路研究中心教授麥爾荀伯格(Viktor Mayer-Schönberger)和經濟學人雜誌資料編輯庫基耶(Kenneth Cukier)共同著作。兩位作者專精電腦及資料科學,對其發展趨勢有獨到的洞見。本書共十章,可分為三個部分:前四章簡介巨量資料及面對巨量資料的新思維;第五章到第八章討論使巨量資料大放異彩的重要關鍵,掌握應用先機;最後三章提醒讀者,巨量資料像一把雙面刃,妥善運用才能讓世界運作的更好,而不是對彼此造成傷害。

一、巨量資料的新思維：大量、雜亂、相關

面對龐大的數據,作者醍醐灌頂三個重要的新思維:(一)樣本等於母體、(二)包容不精確的資料,以及(三)放下對因果關係的執著。這三項新思維顛覆以往小量資料(small data)只能處理特定議題的觀念。

(一)樣本等於母體

以往蒐集資料的難度較高,難以一覽母體的全貌,必須透過隨機取樣取出少數樣本,再以統計方法推估母體的性質。現今獲取資料的難度大幅下降,巨量資料量幾乎就是母體本身,因此得以直接檢視母體的狀態。對巨量資料的思維而言,致力於資料的累積,就越有機會看見全貌。

1 為了與小量資料對應,本文將 big data 譯為巨量資料,但書名仍採用原譯「大數據」。

2 位元組(byte)是電腦資訊的計量單位,每 10^3 為一個計量級,依序為:kilobyte(KB)為千位元組、megabyte(MB)為百萬位元組、gigabyte(GB)為十億位元組、terabyte(TB)為兆位元組、petabyte(PB)為千兆位元組、exabyte(EB)為百京位元組、zetabyte(ZB)為十垓(音同該)位元組、yotabyte(YB)為秭(音同紫)位元組。「五經算術」:「按黃帝為法,數有十等。及其用也,乃有三焉。十等者,謂億、兆、京、垓、秭、穰、溝、澗、正、載也」。

(二) 包容不精確的資料

在小量資料的世界裡，精確度相當重要，因為誤差會在推估過程被放大而使結果失真。巨量資料重視的則是4V：大量(volume)、速度(velocity)、多樣(variety)及真實(veracity)。雖然廣納不同來源的資料，雜亂度就會增加，誤差值也會越多。但是，巨量資料就是在微觀層面犧牲精確度，換取宏觀層面上的新資訊與收穫。對巨量資料而言，2加2等於3.9就已經夠好了(Hopkins and Evelson 2011)。這個世界本來就不是整齊齊地，完整雜亂的資料比規模小的精確資料更貼近現實。

(三) 放下對因果關係的執著

無風不起浪，事出必有因。但是，因果關係不易釐清，武斷的判斷容易導致錯誤的結論。相對地，探討相關關係反而輕易許多，顯著的相關關係代表極高的可能性。對巨量資料而言，極高的可能性就隱含新的見解。舉例來說，「一萬名愛滋病患中，99%的病患在喝下可樂之後痊癒，你會不會讓下一個病患喝可樂？即使不知道原因？」，此時最急迫的事情是拯救病患，趕緊依照「正是如此」去做，暫且不需要太計較「為何如此」，即便無法充分瞭解因果關係，也要看見確實存在的相關關係。

二、累積資料，洞察先機

資料化(datafication)是指將現象以各種方式記錄並保存，史前時代的野牛畫像，就是將現象資料化。隨著文字與數字的發明，人類有更多方式記錄現象；數學的發展，更讓現象得以用數值記錄和運算。資料化與數位化(digitalization)不同，數位化是將資料轉換成二進位供電腦判讀的形式。資料化需要的是察覺重要資訊的洞見和眼光。即使是從未當作資訊的細微訊息，例如網路使用者的點閱動態、麻雀(*Passer montanus*)停留的位置、高鐵行駛的振動狀況等等，都可能蘊含重要的資訊。物理學家von Baeyer(2005)所

說：「萬物的基礎不是原子，而是資訊」。

將資訊資料化之後，接著要提高資料的「選項價值」，包括：重複使用資料、重新組合資料，以及讓資料買一送一。古典的想法認為「特定資料只有特定的用途」，這是在窄化資料的價值。資料價值的環節包括資料、技術和思維，也就是揉合資料持有人、資料專家和巨量資料思維者的觀點，做為處理資料的基本態度。

激發巨量資料蘊含的潛力及價值，要能廣納不同來源的資料，且具備洞察重要資訊的能力。如果未能即時將現象資料化，資訊就難以挽回，甚至永遠消失。因此，必須盡力蒐集各種現象的資料，建立龐大的資料寶庫，才有機會找出藏鏡人，先發制人，掌握先機。

三、莫忘謙卑與人性

「水則載舟、水則覆舟」，巨量資料不是萬靈丹，同樣也會有黑暗的一面。主要三大威脅是：對隱私權的侵犯、根據習性的懲罰(penalty based on propensity)，以及資料獨裁(dictatorship of data)。

諸如通訊、交通、醫療和金融資料，勢必包含私人的通聯、行動、病歷和資金紀錄等隱私資訊。隨著巨量資料的發展，現有的隱私權相關法律可能不足以規範巨量資料對隱私權的衝擊，必須適當修改法律保障個人隱私。

「預測」是巨量資料令人躍躍欲試的功能，如消費行為和運動員的表現，同樣地，也能應用於犯罪行為。但是，即便預測結果顯示某人未來犯罪機率很高，也不能在犯罪未發生之下逕行逮捕。貿然的運用巨量資料，很容易對特定特質的人產生歧視與傷害。

資料獨裁是指過度依賴資料盲目思想。數字不會說話，讓資料說話的是人，解讀資料的也是人。無論資料規模大小，還是在人的手上發揮。巨量資料只是工具的一種，可以增進世界的福祉，也能摧毀世界。面

對巨量資料的龐大能量，終究要回歸人性，秉持審慎與謙卑態度學會掌控資料，而非被資料掌控。

見解評述

一、取樣會產生什麼問題？

隨著資料管理的軟硬體如雨後春筍般地出現，智慧型手機和平板電腦的普及化，以及多樣的行動應用程式(mobile application, APP)，大幅增加資料管理的方法。再加上資料傳遞方式不再受限於有形的線路，無線網路及第四代行動通訊技術標準(the fourth generation of mobile phone mobile communication technology standards, 4G)使通訊更加快速便利。資料科學家發現，當資料多到一定程度時，量變產生質變，資料的應用與展現都與以往大不相同，巨量資訊的時代早已展開。

巨量資料廣為流傳之後，有些人熱情擁抱，有些人批判缺點(Crawford 2013)。主要是巨量資料與統計學的觀點衝突，一方認為巨量資料足以取代統計學，另一方則認為巨量資料內的誤差使其毫無價值。事實

上，兩者之間並無衝突，端看如何交互運用，並參透資料裡蘊含的訊息。

取樣(sampling)是難以蒐齊母體資料時所採用的方法。要能有效推估母體，關鍵是落實隨機取樣(random sampling)，但是，真正的隨機並不容易落實，隨機的定義是「所有的樣本被選取的機率相等(Quinn and Keough 2002)」。因此，幾乎只有透過電腦程式才能落實真正的隨機。例如抽籤就不是完美的隨機取樣，因為抽籤者對籤筒「中間」和「邊緣」的籤支可能有不同的偏好，使兩者被抽到的機率不相等。因此，巨量資料就是要將母體資料一網打盡，讓樣本就是母體本身。

有些狀況可以使樣本完全等於母體，例如全國高中生的身高，但有些則難以達成，只能盡力讓樣本逼近於母體，例如全臺灣鳥類的總數量。即便如此，巨量資料還有機會逼近或等於母體，無論取樣設計再縝密，最終結果依舊是樣本的估計值，估計值永遠不會等於母體本身。

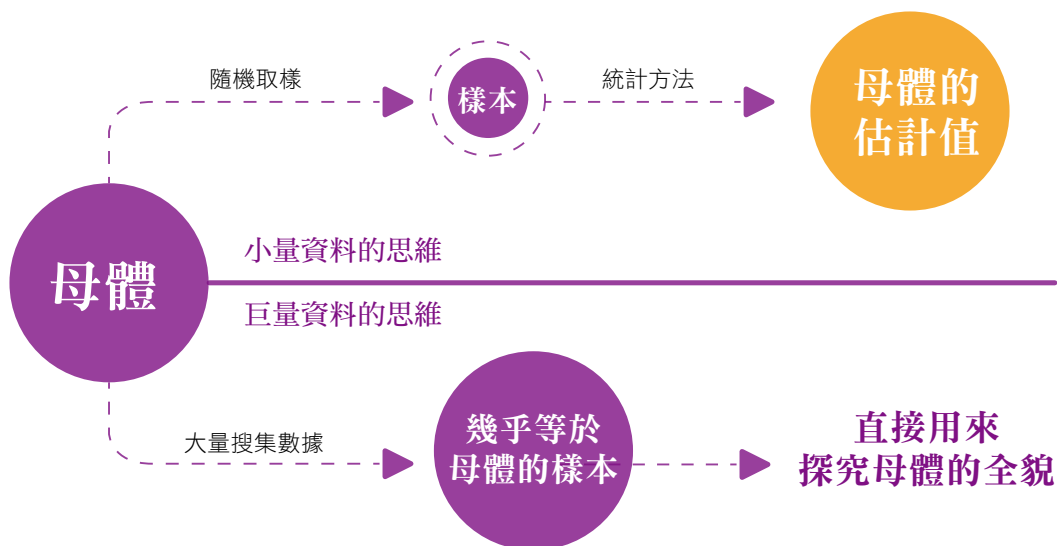


圖 1. 小量資料與巨量資料對「樣本」概念的差異。

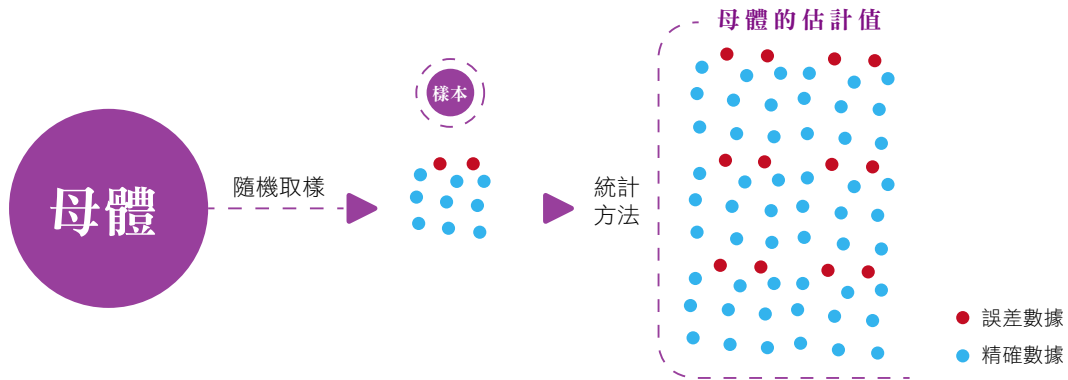


圖 2. 統計方法容易將樣本中的誤差放大，因此特別要求精確。

二、資料不精確沒關係嗎？

統計學的精髓在於以少估全的技術，以少數樣本推估母體的全貌。精確值在統計中之所以重要，是因為樣本會被統計方法放大成母體估計值，誤差值則同時被放大，意即「差若毫釐，謬以千里」。

巨量資料廣納各種來源的資料，其中所含的偏差值，是常常被提及的缺點。此時應該討論巨量資料和小量資料所適用的精確度。如圖3所示，一般認為巨量資料的品質參差不齊，小量資料的精確度較高。因為小量資料，來自審慎的試驗設計、於固定的操作方法、相同的測量者及儀器，儘可能排除各種產生誤差及影響重要變數的因素。巨量資料沒有經過這些操作及控制，使誤差值較多。

然而，實際的狀況如圖4所示。小量資料雖然經過謹慎的測量，但是各種現象之間的交互作用極其複雜，實務上很難操控所有的變因，例如在物理實驗中，要物體落下時完全排除空氣阻力或物體移動時完全排除摩擦力的影響就相當困難。再加上不確定原理 (uncertainty principle)(Heisenberg 1927)的效應³，應該要想清楚問題適用的精確度為何，2加2必須等於4.00還是3.9？而不是一味的追求沒有必要的精確。如果要

回答「臺灣的海岸線有多長？」這個問題，分別拿最小單位為1m、1cm和1mm的尺去量，不僅所得到的數值差距甚大，其意義也南轅北轍。巨量資料捨棄對精確度的追求，從中擷取可相互契合的資料。在大量蒐集資料之後，如圖5所示，可運用的資料同樣也會增加。精確的資料著實令人放心，但是以管窺天仍舊難以掌握全貌。

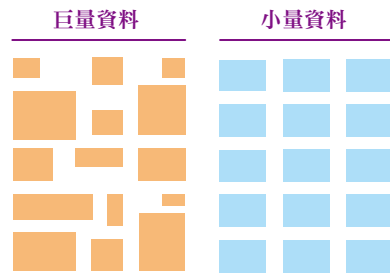


圖 3. 一般對於巨量資料與小量資料精確度的想像。

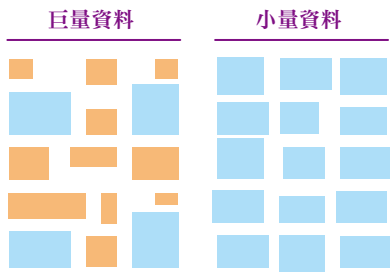


圖 4. 巨量資料與小量資料的實際精確度，藍色矩形指巨量資料中可契合的資料。

3 曾經被稱為「測不準原理」，德國物理學家維爾納·海森堡 (Werner Heisenberg) 指出，測量動作不可避免地攪擾了被測量粒子的運動狀態，因此產生不確定性，從極細微的尺度來看，測量準確這件事不可能發生。參閱 Heisenberg, W. 1927. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. Zeitschrift für Physik, 43 (3-4): 172-198.

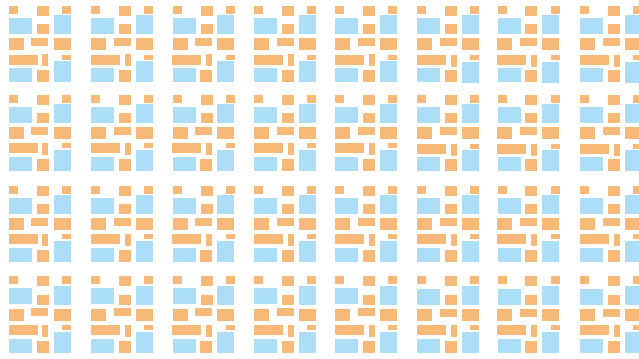


圖 5. 大量資料蒐集累積之後，巨量資料中可運用的資料也會大幅增加。(林大利 繪)

三、相關關係與因果關係

當現象A引起現象B時，現象A稱為「因」，現象B則稱為「果」。如果A與B之間還有一連串的因果關係，這個因果關係鏈則稱為「機制」。探討因果關係之前，通常先檢視兩者之間是否具有相關關係。相關關係是指兩個現象之間，其中一個發生時，另一個也會跟著發生的關係。因果關係會造就顯著的相關關係，所以相關越顯著，就越令人相信這是重要的因素，但是，相關關係並不等於因果關係(Tufte 2006)。

相關關係不等於因果關係的原因有幾個：(一)兩個現象共同發生可能僅是純屬巧合；(二)可能倒果為因，或是兩者互為因果；(三)另一個因素對兩個現象共同作用；(四)還有其他因素一起對結果作用(Pearl 2000)。更重要的是，許多現象與現象之間的影响，以及彼此之間的交互作用相當複雜，不容易抽絲剝繭加以釐清，甚至永遠無法確切地弄清楚整個因果關係的結構。

美國心理學家丹尼爾·康納曼(Daniel Kahneman)認為大腦的思考系統可以分為兩大類，一個是成本低，必須在幾秒鐘快速下結論的「快思系統」，另一個是成本高，需要多加琢磨思考的「慢想系統」(Kahneman 2011)。大腦習慣採取耗費能量較低的策略，讓思考容易傾向因果關係，因此輕易地建立因果關係，但是也容易誤解事實。

全球目前面臨許多急迫的危機，包括氣候變遷、資源不均、物種滅絕和傳染病擴散。面對這些緊急狀態，已經沒有時間鉅細靡遺的探究因果關係，如果能以相關關係發現可能的解決之道，那就應該趕緊嘗試。「大數據」並非主張因果關係不重要，只是暫時不將它當主角，先解決急迫的問題要緊。

生物多樣性保育與巨量資料共舞

我開始賞鳥至今已經13年，把鳥納入望遠鏡的視野中、正確地辨識種類、學習鳥類的知識。直到現

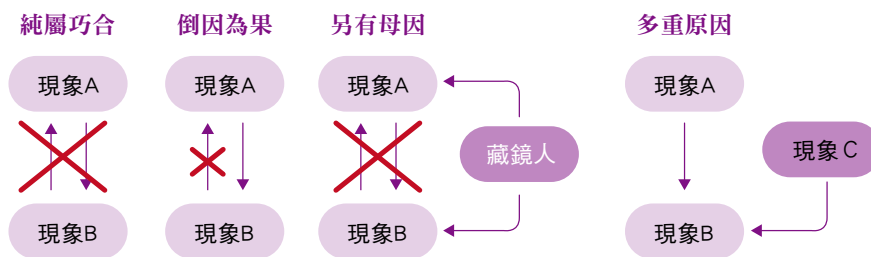


圖 6. 相關關係不等於因果關係的幾個可能原因。



僅需要將生物照片、發現時間及地點上傳，經鑑定後儲存於資料庫，即可成為有效的資料。

在，在行政院農業委員會特有生物研究保育中心(以下簡稱特生中心)服務，成為以鳥類為研究對象的研究者、以賞鳥為休閒活動的自然愛好者。鳥類，早已成為我生活中的許多部分。

我從賞鳥到研究鳥類，不斷地在問幾個問題：「哪裡鳥多？哪裡鳥少？為何而多？為何而少？如何增多？如何減少？」在我的求學歷程中，以「假說—演繹模型(hypothetico-deductive model)」⁴做為主要的科學方法，針對特定的問題提出合理的假說，並設計蒐集數據的實驗，接著分析數據、解釋結果。限於人力、經費與時間，研究主題所涵蓋的空間與時間範疇不大，能探討的議題也有限，但是對研究生而言，已經足以探索新知。

進入公部門之後，我的思維和研究方向開始轉變。特生中心是隸屬行政院農業委員會的生物多樣性研究機構，必須將研究及探討的問題提升至國家層級。國家所涵蓋的空間面與時間面向，比我以往的經驗大上許多。為了有效的掌握我國生物多樣性的現況及變化，我開始嘗試以公民科學(citizen science)⁵的方式執行，徵求志工在全國各地記錄生物的時空資訊。雖然更快速且大量地蒐集資料，但是也蘊含許多誤

差與偏差，無法輕易地以統計技術處理。面對這個問題，苦思許久仍不得其解，直到「大數據」指引新的思維，讓我知道下一步怎麼走。

一、生物多樣性指標

2010年，聯合國生物多樣性公約(Convention on Biological Diversity, CBD)秘書處於日本愛知縣召開第十屆大會，檢視各締約國執行「2010生物多樣性目標(2010 Biodiversity Targets)」的成效。會後發表的「全球生物多樣性展望第三版(Global Biodiversity Outlook 3, GBO3)」指出，大多數締約國並未達成目標(Secretariat of the Convention on Biological Diversity 2010)。更進一步指出：「生物多樣性狀態」的劣化趨勢沒有顯著減緩；「生物多樣性面臨的壓力」則呈現上升趨勢(Butchart *et al.* 2010)。雖然國際上對於生物多樣性保育的態度越來越積極，逐漸可見成效，但是，生物多樣性流失並未減緩，各國仍需持續努力。為此，秘書處重新設置更嚴格的「愛知生物多樣性目標(Aichi Biodiversity Targets)」要求各締約國至遲於2020年之前達成(Secretariat of the Convention on Biological Diversity 2010)。

為了清楚瞭解生物多樣性的保育成效，秘書處與聯合國環境規劃署結合各國相關組織，組成「生物多樣性指標夥伴關係(Biodiversity Indicators Partnership, BIP)」。針對愛知生物多樣性目標的內容，設計適當的

4 意指羅列合乎邏輯的虛無假說(null hypothesis)及替代假說(alternative hypothesis)，再以統計學執行假說檢定，以否認原則排除虛無假說，接受替代假說的科學方法。

5 透過民眾參與執行的科學研究方法，通常適用於大範圍及長時間追蹤的研究。

指標，例如森林覆蓋的百分比和物種的相對族群量等等，以檢驗目標執行進度。我國雖然並非屬於生物多樣性公約的會員國，但是仍然應盡力達成大會的要求，積極達成目標所訂立的標的。

二、公民科學：公民參與的科學

瞭解國家級的生物多樣性現況及變化，需要大空間尺度及長期監測，必須蒐集大量的資料，且能有效的整合與運用(Hochachka *et al.* 2012)。這些資料的取得需耗費大量人力與資金，不容易由研究人員獨立完成，於是民眾共同參與的「公民科學」便應運而生(Dickson and Bonney 2012)。公民科學是指公民與科學家合作，由民眾擔任主要的資料蒐集者，科學家設計方法和分析數據，探討共同關注的議題。公民科學快速蒐集大尺度及同步性高的巨量資料，目前已經應用於天文學、環境監測及生物多樣性保育。

公民科學資料的性質與巨量資料相當相似，廣納來自全國各地的民眾參與，大幅提升資料量，但是資料的結構與性質也比試驗設計法大不相同。累積生物時空分布資料的公民科學，資料組成包括生物名稱、時間資訊與空間資訊三者。這一類計畫，通常採用Web 2.0⁶的概念建立資料庫，讓使用者主動上傳生物分布資訊。智慧型手機和平板電腦的普及，使人人隨身攜帶相機和全球定位系統(Global Positioning System, GPS)。

藉由這個趨勢，特生中心研究人員運用臉書(facebook)社團做為資料輸入平台：民眾在野外發現生物，用行動裝置拍照、打卡，將照片及時間地點等資訊上傳至社團頁面，適用於小型動物和因車禍死亡



透過公民科學可快速展現鳥類路殺的分布熱點，不同深淺綠色為海拔高度，越深表示海拔越高。黃色為保護區分布範圍。

的動物，民眾不一定要會辨認生物，由研究人員鑑定即可。因此，上手容易、操作簡單的公民科學活動，便能快速累積資訊。目前已經在運作的團隊包括：「四處爬爬走：路殺社」、「慕光之城：蛾類世界」、「蝸蝸園：臺灣陸生蝸牛交流園地」、「蛛社會社：臺灣蜘蛛研究」，以及「台灣植物資料調查及植物物候觀察記錄團」。

隨著電腦演算法與統計方法的進展，以及生物分布資料快速累積，越能建立良好的「物種分布模式」。物種分布模式是運用演算法所建立的生物分布與環境關係的量化工具(Guisan and Zimmerman 2000)。通常與地理資訊系統揉合，將生物與環境的圖層套疊成生物分布機率圖，與生態及演化的觀點揉合後，就能探討生物分布的機制及預測未來的分布趨勢(Elith and Leathwick 2009)。目前且已廣泛應用於氣候變遷對生物的影響、入侵物種擴張，以及瀕危物種的棲地管理。

⁶ 是指由使用者主導而生成的網站內容的網際網路產品模式，著重網站管理員與瀏覽者的互動。

三、開放吧！資料

隨著全球生物分布資料逐漸增加，為了有效整合資訊，聯合國透過生物多樣性公約制定全球生物多樣性資訊的交換機制，於2001年成立「全球生物多樣性資訊機構」(Global Biodiversity Information Facility, GBIF)。GBIF的主要任務是彙整全球的生物多樣性資訊，資料完全開放對全世界分享，任何人皆可自由下載運用。

GBIF的運作方式，是與全世界各資料管理組織結盟，形成夥伴關係。同時，GBIF邀請參與國擔任GBIF資料的節點，例如歐洲生物多樣性資訊網絡(European Network of Biodiversity Information, ENBI)就是歐洲的節點。透過節點的串聯，各國會將資料匯集到節點，再彙整到GBIF，同時也做為國家之間聯繫的窗口。截至2014年，GBIF已經包括609個資料管理組織、15,151個資料集、145萬種生物，共累積4.4億筆生物分布資訊，實屬全世界最大的生物巨量資料庫。

臺灣也不落人後，自2003年起由中央研究院生物多樣性研究中心邵廣昭博士統籌，成立臺灣生物多樣性資訊機構(Taiwan Biodiversity Information Facility, TaiBIF)，做為GBIF的臺灣節點，目前共累積250萬筆的生物分布資訊。公民科學所蒐集的生物分布資料，也一致匯入TaiBIF資料庫，最終進入GBIF與全球分享。

生物分布的資料相對單純許多，除了珍貴稀有物種和走私犯覬覦的物種之外，大多數生物的分布資料都鮮少涉及隱私權問題。因此，開放分享生物分布的資料，適合做為「開放政府資料(open governmental data)」政策的初步嘗試。開放政府資料是指將政府持有的大量資料數位化之後，不限定形式、即時公開且無須授權的開放於網際網路。基本緣由在於政府取得及應用資料，皆以稅金為主要資金，政府依其職責代替全民蒐集和管理。韓非子之十過篇：「聖人之治，藏於民，不藏於府庫」。資料的自由開放，等同是讓更

多民眾參與，透過市場競爭與集思廣益，提升資料的應用價值。

魏書列傳第五十六篇，甄琛：「善藏者藏於民，不善藏者藏於府。藏於民者民欣而君富，藏於府者國怨而民貧」。資料就像金錢一樣：被花掉的錢才是錢，否則就只是遺產；被應用的資料才稱為資料，否則只是在占據記憶體容量。讓更多人有機會接觸資料，就越有機會提高資料的應用價值。

四、分析巨量資料的挑戰

處理巨量資料，必須清楚瞭解資料的特性。以公民科學的資料為例，包括參與者辨識與調查能力不盡相同對資料結構造成的偏差(Cooper *et al.* 2007; Hochachka *et al.* 2007; Tulloch 2013)、不完美的偵測度(imperfect detection)導致「錯誤缺席(false absence)」的資料(McClintock *et al.* 2010)、資料的時空分布不均(Boakes *et al.* 2010)，以及多數資料來自少數使用者的現象(Sullivan *et al.* 2014)，這樣的資料特性稱為不完美資料(imperfect data)(Foody 2011)。由於不完整資料難以直接套用至統計分析，結果都不甚完善。讀過「大數據」之後，瞭解這是巨量資料的正常現象，重點在於巧妙的從中擷取精華，挖掘資訊，再進入分析階段。

雖然公民科學計畫所產出的資料或多或少具有偏差與缺點，但是資料庫所累積的資料量遠遠超過以往田野調查的資料，大量的資料之下仍然藏有重要分布資訊與變化趨勢。為了有效應用所內含的分布資訊，應該探討如何選擇適當的資料應用、偏差校正、資料擷取等方法，使其結果能有客觀的解釋力，提升資料的應用價值，以解決環境及保育議題。

結語：資料不缺訊息，只缺發現

一個世紀以前，因為資料有限，發展統計學來參透世界的全貌，至今發展更多精巧嚴密的統計方法。

科技的發展讓資料的紀錄、儲存與傳遞變得更加快速便利，造就超乎想像的巨量資料。於是，資料科學家發現資料的量變產生質變，且能以新的方法來探索訊息。

巨量資料初試啼聲之後，部分人宣稱巨量資料足以揚棄統計學，認為不應該只思考如何讓馬跑得更快，而是發明汽車。這樣的宣稱造成巨量資料與統計學雙方的對立。巨量資料是否會引起下一波資訊革命，還有待觀察。巨量資料確實是新的曙光，但是尚未發展成熟，無論選擇大膽嘗試或且戰且走，都應該以謙卑的態度探索知識。

科學是人類瞭解世界的系統性活動，探索知識就是在認識世界。許多現象的交互作用頻繁又複雜，彼此之間互依互存。巨量資料與小量資料的思維並不衝突，而是相輔相成，從巨量資料中擷取精華，再運用統計知識加以分析闡述。牛頓說：「我像是在海灘上玩耍的孩子，一會兒發現美麗的貝殼，一會兒發現特別的石子。但是，對於真理所在的茫茫大海，我卻一無所知(Brewster 1855)」。巨量資料也如同浩瀚的汪洋，它不缺少知識與資訊，只是缺少發現。

相關網址

四處爬爬走：路殺社

<https://www.facebook.com/groups/roadkilled>

臺灣野生動物路死觀察網

<http://roadkill.tw/>

慕光之城：蛾類世界

<https://www.facebook.com/groups/enjoymoths/?fref=ts>

蝸蝸園：臺灣陸生蝸牛交流園地

<https://www.facebook.com/groups/283177105146997/?fref=ts>

株式會社：臺灣蜘蛛研究

<https://www.facebook.com/groups/SpiderTw/?fref=ts>

台灣植物資料調查及植物物候觀察記錄團

<https://www.facebook.com/groups/409531105787849/>

全球生物多樣性資訊機構

GBIF <http://www.gbif.org/>

臺灣生物多樣性資訊機構

TaiBIF <http://taibif.tw/>

歐洲生物多樣性資訊網絡

ENBI <http://www.enbi.info/>

生物多樣性指標夥伴關係 BIP

<http://www.bipindicators.net>



有效率的長期監測需要全民共同參與。