如何於生物多樣性監測中提升公民科學資料的品質

How to improve the data quality in citizen science in biodiversity monitoring?

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

公民科學的興起

2010年，聯合國生物多樣性公約(Convention on Biological Diversity, CBD)秘書處訂立新的生物多樣性十年目標「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。各締約國為了瞭解國家生物多樣性現況與變化，設計許多生物多樣性指標(biodiversity indicators)，做為評估保育策略和規劃經營管理計畫的觀察客體。

為瞭解大範圍的生物多樣性現況及變化趨勢，必須在大時空尺度下長期監測指標的變化，蒐集大量的資料，並且有效的整合與運用。以物種保育為例，瞭解該物種的分布、相對族群量、棲地偏好等，是規劃保育及經營管理策略的基礎資訊(Gorrod and Keith 2009)。專業調查員所做調查，能在較小的時空尺度之下獲得精確的時空分布資料，但是卻無法快速蒐集大範圍的分布資料，而且須耗費龐大的人力與經費。因此，由政府機關、學術組織、

行政院農業委員會特有生物研究保育中心自2009年起，已推動諸多公民科學計畫，圖為2016年路殺社年度大會合影。(林沛晨 攝)

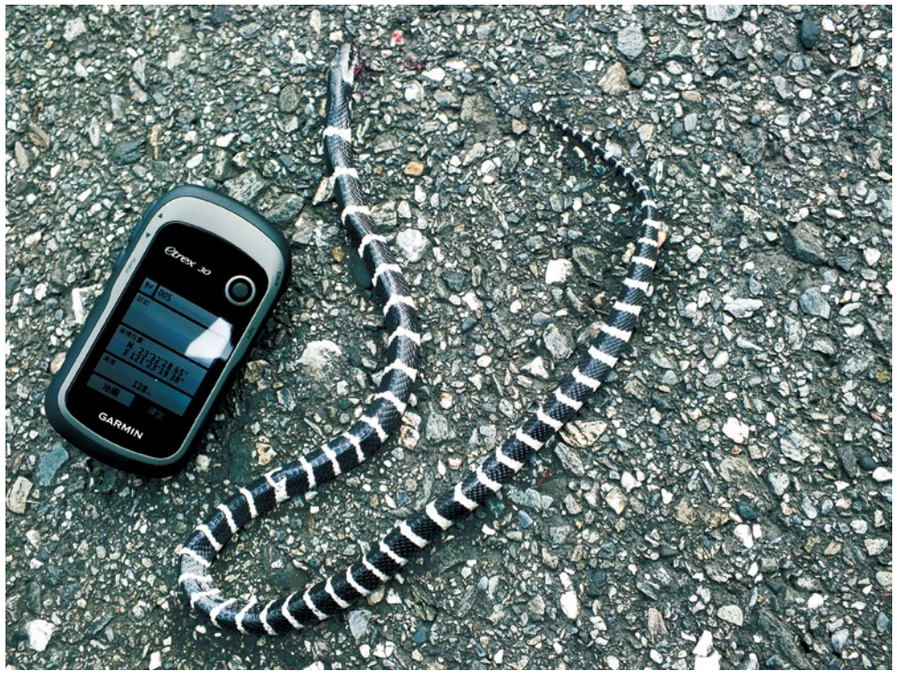
臺灣的政府機關、大專院校及民間組織已合作執行許多公民科學計畫。

民間社團及業餘愛好者所共同合作的「公民科學(citizen science)」計畫便應運而生。由各地民眾參與的公民科學計畫，能在較大的時空尺度之下獲得大量的生物分布資料，彌補傳統由少數人調查的弱點(Dickinson *et al.* 2012a)。研究人員為志工量身打造適合的長期監測方法，由參與民眾蒐集可供科學分析的數據資料，再由研究人員進一步分析探討(Cooper *et al.* 2007; Bonney *et al.* 2009；Sullivan *et al.* 2009；Dickinson *et al.* 2010； 2012a；2012b)。

以公民科學方式監測自然現象，能快速獲得大量、廣時空尺度及同步性高的資料，顯示大範圍的生物多樣性現況及變化趨勢(Kelling *et al.* 2009；Hochachka *et al.* 2010；Hochachka *et al.* 2012)，亦能提升參與者的科學知識與素養，兼具科學研究及推廣教育的價值(Bonney *et al.* 2009；Dickinson *et al.* 2010)。近年歐洲的生物監測計畫中，將近86%是與自願參與的志工合作完成，大幅節省聘用專業人員的費用，以及交通所需的金錢與時間成本，每年約節省400萬歐元(Schmeller *et al.* 2009；Levrel *et al.* 2010)，顯示廣時空尺度的監測對公民科學的需求逐漸提升。以公民科學為機制所執行的計畫已經應用於天文學(如星系動物園Galaxy Zoo http://www.galaxyzoo.org)、衛星雲圖判讀(如氣旋中心計畫Cyclone Center Project http://www.cyclonecenter.org)、環境監測及生物多樣性監測，如奧杜邦鳥類學會的聖誕節鳥類調查(Christmas Bird Count, CBC)。

公民科學的類型大致可分為兩類：「系統型計畫(systematic scheme)」和「機會型計畫(opportunistic scheme)」(Lewendowski and Spech 2015)。系統型計畫是指訂立嚴謹標準作業程序的公民科學計畫，在資料蒐集階段就先規劃，將志工不同的調查能力和努力量標準化，所獲得的資料偏差較低，較容易以傳統的統計分析，例如英國、美國及臺灣的「繁殖鳥類調查(Breeding Bird Survey，BBS https://goo.gl/dPQI2a)」和「臺灣新年數鳥嘉年華(Taiwan New Year Bird Count, NYBC Taiwan http://nybc.bird.org.tw/)」。此類的公民科學計畫，參加門檻較高，參與者必須接受培訓課程的訓練，熟悉監測方法及流程，才能實際蒐集資料(Dickinson *et al.* 2010)。以「臺灣鳥類生產力與存活率監測(The Monitoring Avian Productivity and Survivorship program, Taiwan; MAPS Taiwan https://goo.gl/HHrJ5u)」為例，必須透過鳥類繫放檢視鳥類的年齡與性別特徵，才能探討鳥類的族群結構(population structure也就是鳥類的人口金字塔)。鳥類繫放是相當困難的技術，若技術不夠純熟，很容易造成鳥類受傷甚至死亡。因此，志工在參加之前，必須先接受過相當嚴格的訓練，才能實際參與計畫。

「機會型計畫」之目的在於累積大量的生物時空分布資料，尤其物種、時間及空間三者，能快速累積生物的分布資料，使資料越來越容易取得(Guisan *et al.* 2006；Graham *et al.* 2008)。任何使用者都能夠在資料庫或平臺輸入所需資料，例如累積賞鳥紀錄的eBird資料庫(Sullivan *et al.* 2014 http://ebird.org/content/ebird/)。機會型計畫的參與門檻相對較低，使用者可以隨時提供資料庫所需的資訊，甚至只須要上傳照片及時空資訊到網路平臺或資料庫中，由專業研究人員鑑定物種即可，使用者不一定要具備鑑定能力。機會型計畫就很適合運用在蛾類、蝸牛、蜘蛛等容易觀察拍攝的小型無脊椎動物，或是死亡動物的監測。臺灣的「四處爬爬走：路殺社」、「臺灣野生動物路



機會型計畫的「四處爬爬走：路殺社」，隨時皆可接收路死動物的時空資訊。

死觀察網，https://roadkill.tw」和「慕光之城：蛾類世界之飛蛾資訊分享站 http://goo.gl/su4eyX」都是機會型計畫的典型案例。

公民科學資料的特性

公民科學資料承載大量且豐富的資訊，一般而言，可以探討相對族群量的變化趨勢、生物的棲地利用與偏好，以及建立物種分布模型(Species Distribution Model, SDM)。然而，在試驗設計與資料分析方面也須面臨前所未有的挑戰(Hochachka *et al.* 2010；Hochachka *et al.* 2012；Lewendowski and Spech 2015)。首先，經營公民科學計畫不僅需要合適的資料管理軟體與硬體，在研究規劃方面，也必須設計適合參與民眾的調查方法，以維持其參與動機及熱情(Hochachka *et al.* 2012)。第二，必須清楚瞭解資料特性並掌握資料品質，包括參與民眾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)、不完美的偵測度(imperfect detection)導致「錯誤缺席(false absence)」的資料 (McClintock *et al.* 2010)，以及資料的時空分布不平均(Boakes *et al.* 2010)。第三，大時空尺度的生態現象不容易適當的模式化，因為各個時空層級皆有不同的因素影響生物的時空分布，包括遷徙、播遷、競爭、及獵食等生物之間的交互作用(Scott *et al.* 2002)。

公民科學資料庫的資料結構取決於使用者的組成，由於資料來自許多志工，不同來源的資料自然會有較高的變異與偏差。此外，物種偵測能力的差異(MacKenzie 2005；Guisan *et al.* 2006；Hartle *et al.* 2009)、回傳資料不完整(Conlisk *et al.* 2009)、空間資訊模糊(Freeman and Moisen 2008；Johnson and Gillingham 2008)，以及生物存在與否的資料比例不均(Real *et al.* 2006)，都是常見的現象，這樣的資料特性稱為不完整資料(imperfect data; MacKenzie 2005；Royle *et al.* 2007；Foody 2011)。公民科學資料因而很容易受到質疑，甚至被排除在分析之外(Engel and Voshell 2002；Foster-Smith and Evans 2003)。傳統的統計方法是以樣本推估母體為主要原則，這樣的方式會放大變異與偏差，因而特別講究資料的精確度和準確度，盡可能排除產生誤差的因素，主要分析來自審慎試驗設計的資料(Mayer-Schönberger and Cukier 2014)。公民科學資料對傳統的統計分析而言，過高的變異與偏差無法真實反映生態系統的實際狀況(Gorrod and Keith 2009)。

辦理培訓課程是公民科學的重要工作。

特生中心曾於2012年舉辦公民科學特展，介紹世界及臺灣各地的公民科學計畫。(施禮正 攝)

因應公民科學機制所造成的資料偏差，研究人員須建立適合參與民眾操作的標準作業程序，並在執行調查之前辦理訓練課程及說明會，讓參與者瞭解目的與操作方法，以降低操作不當所致的偏差及資料中的雜訊(Bonney *et al.* 2009)。

時空因素方面，常可見資料分布不均的現象(Hochachka *et al.* 2012；Sullivan *et al.* 2014)，這些訊息可能是生物分布的實際狀況，也可能是其他因素所造成。以路殺社及臺灣新年數鳥嘉年華的資料為例，臺北地區因人口密度較高，其資料分布點和樣區圓的密度相當高，但臺北市可能並非真正的生物多樣性的熱點，只是居住於北部的志工較多所致。全球的資料庫也有類似的現象，如GBIF及eBird資料庫中，亞洲國家可能因語言隔閡或國家政策，導致資料來源較少。另一方面，時空資訊不夠精確也是常見的問題，例如以較大的行政區域名稱描述調查地點如「花蓮縣秀林鄉」或「太魯閣」，或以公路名稱如「中橫公路」或「大雪山林道」描述調查路線，皆無法讓資料使用者得知確切的調查位置，若未妥當處理，可能導致空間資訊錯誤(Freeman and Moisen 2008；Graham *et al.* 2008；Johnson and Gillingham 2008)。

提高公民科學資料品質

雖然公民科學資料或多或少具有偏差與缺點，但是累積的資料量遠遠超過傳統田野調查的資料，大量的資料之下藏有重要分布資訊與族群變化趨勢。處理公民科學資料可由下列四

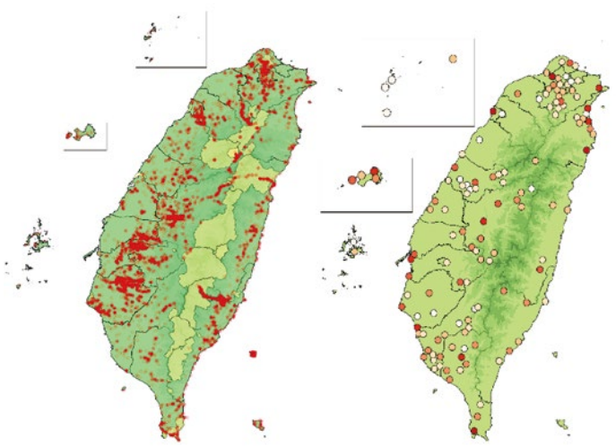
個角度著手：(一)個人資料蒐集面：透過訓練課程與現地示範，使個人的觀察資料更準確，包括察覺生物、物種辨識和族群量估算的能力。(二)標準化取樣：為監測方法訂立詳細且嚴謹的標準作業程序，使志工們在不同時空之下，運用一致的方法調查生物，可有效不同調查志工之間降低偏差、避免主觀判斷(Crall *et al.* 2011)。(三)時空重複取樣：在不同樣點及時間點之間由不同志工重複調查。由於每位志工可能各自有不擅長偵測的物種，因此，由不同志工重複調查相同樣區，可以互相彌補容易忽略的物種，降低因偵測度所造成的影響(Lovett *et al.* 2007)，同時也能提高樣本數。(四)樣本數：使樣本數增加足以進行統計分析，甚至可經由資料挖掘(data mining)篩選分析所需要的資料(Yoccoz *et al.* 2001；Legg and Nagy 2006；Lovett *et al.* 2007；Dickinson *et al.* 2010)。

在執行監測之前，進行志工培訓課程，是提升資料品質的常見方法，也是執行公民科學計畫中不可忽略的重要工作。培訓課程中，先對志工說明監測計畫的目的，是為了瞭解什麼現象？或解決何種問題？使用何種方法執行監測工作？尤其要仔細說明監測方法設計的考量，可使志工瞭解應該注意的關鍵細節，以避免資料產生過高變異與偏差(Lovett *et al.* 2007；Dickinson *et al.* 2010)。物種辨識的培訓，運用或製作辨識物種的輔助工具，如辨識摺頁，可避免物種辨識能力不足或錯誤的問題(Lotz and Allen 2007)。安排田野調查的實際示範與演練，也能有效提高志工的調查能力。

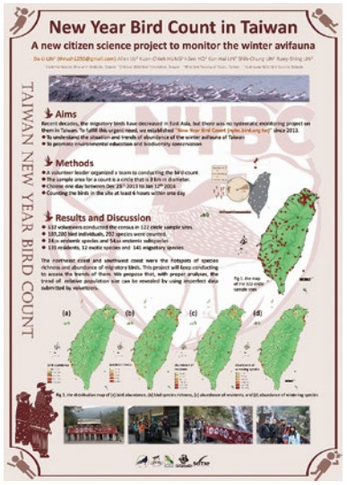
從資料層面而言，志工繳回的資料必須經過審核，再三確認不合理的物種紀錄和過高的族群量。與努力量相關的參數，如調查時間長度、調查範圍面積、調查距離、參與人數等，都必須確實記錄，以此做為資料標準化的主要依據(Butt *et al.* 2013)。統計分析方面，公民科學資料難以直接分析，往往需要透過資料篩選或建立模式等高等統計方法分析。隨著電腦演算法與統計分析方法的發展，以及生物分布資料快速且大量的累積，建立物種分布模型的方法日益增加，例如邏輯複迴歸分析(Logistic Multiple Regression)、廣義線性模型(Generalized Linear Model, GLM)、卜瓦松迴歸(Poisson regression)、廣義附加模型(Generalized Addictive Model, GAM)、結構方程式模型(structural equation modeling)、貝氏統計模式(Bayesian model)、支持向量機(supporting vector machine)、及最大熵演算法(Maximum Entropy, MaxEnt)及生態區位因素分析(Ecological Niche Factor Analysis, ENFA)等(Guisan and Zimmermann 2000；Guisan and Thuiller 2005)。建模方法的選擇是相當關鍵的步驟，各種不同的建模方法各有優點及限制，其所適用的物種特性、環境因子特性及資料結構也不盡相同，不同的建模方法所產出的預測結果也可能完全不同(Elith 2000；Moisen and Frescino 2002；Johnson and Omland 2004；Segurado and Ajaujo 2004)。

經營志工

志工是公民科學的核心之一，也是蒐集資料的主要成員，因此，在規劃公民科學計畫，許多細節都需要為志工考量。參與公民科學的調查志工，常常會有特別的偏好。舉例而言，機會型計畫中，志工對稀有物種的興趣往往比普遍種高出許多，因此會比較容易蒐集稀有物種的資料，而缺乏普遍種的資料(Matteson *et al.* 2012；Sarda-Palomera *et al.* 2012)。然而，這樣的現象反而可以將稀有種的分布資料蒐集的更完善，更真實呈現稀有種的分布狀況。也因此，機會型計畫較適合探討生物時空分布(但仍然會有偏差)；系統型計畫則較適合監測生物



公民科學資料容易呈現時空分布不均的狀況(左)，臺灣新年數鳥嘉年華的樣區圓則以北部最密集，但東部仍相當稀疏(右)。



發布公民科學監測成果，也是讓志工看到資料貢獻的重要方式。

舉行成果發表記者會，能讓志工看到參與成果，也能透過媒體達到宣傳效果。

公民科學的志工都是因關心生物或環境而積極主動參與。

帶領志工到現場示範也是效果理想的培訓方式。
(蘇美如 攝)

族群的變化趨勢(Matteson *et al.* 2012；Sarda-Palomera *et al.* 2012)。

志工參與公民科學的過程，也是一種自由選擇學習(free choice learning)的過程，隨著參與的經驗累積、從培訓課程學習相關知識與技術，志工的偵測及辨識能力都會進步(Kelling *et al.* 2015)。如此，長期參與的志工，其調查成果往往會越來越好。Miller等人(2012)發現，隨著參與培訓的次數越多、經驗越豐富，資料的品質顯著提升。但是，研究者在分析資料時必須審慎留意這個現象，有時候並不是生物的種類或數量有明顯的提升，而是志工的調查能力進步所致。

在理想的狀況下，由同一位志工長期穩定的參與，是維繫資料品質的關鍵之一。因此，如何讓志工願意長期參與，也成為公民科學規劃者的重要課題(Uychiaoco *et al.* 2005；Edgar and Stuart-Smith 2009)。由於志工並非受聘僱而來，而是基於自身的興趣，使其瞭解參與公民科學的重要性與貢獻、能與同好交流競爭、獲得研究者的肯定、回饋與感謝，都是維繫志工參與動機及意願的重要工作(Rotman *et al.* 2012)。如果調查方法設計得太複雜困難，很容易降低志工的參與意願(Delaney *et al.* 2008)。然而，過於簡單的方法反而會降低資料品質，因此公民科學調查方法的規劃，必須在志工與科學方法之間做權衡，既能讓志工容易執行，也能客觀地呈現生物多樣性的狀況(Dickinson *et al.* 2010)。

公民科學的特色在於納入許多志工參與，使資料的質與量產生相當大的變化，與試驗設計產生的資料大相逕庭。要維繫公民科學資料的品質，瞭解志工的個人特質與能力，多從人性的角度規劃公民科學計畫，便是重要的思考層面。在經營公民科學計畫的過程中，絕對不能輕易地運用金錢做為犒賞，否則很容易扭曲志工的貢獻價值。近幾年來，公民科學如雨後春筍般快速出現，累積且開放了相當大的生物分布資料(GBIF 2015)。在公民科學推展初期，從各種層面著手，儘可能將資料品質提升。未來運行穩定之後，公民科學資料將會是保育和經營管理的重要資料來源，成為反映生物多樣性現況的重要基礎資訊。