

20年磨一劍 — 讓生態資料與世界接軌

吳佳奇、何芷蔚、端木茂甯

生物多樣性資料共享的開始

在大數據概念還未普及時，聯合國經濟合作暨發展組織（OECD）就已經針對生物多樣性的資料提出建議：「國際間需要建立可以自由存取生物多樣性資料與資訊的機制，以藉由提供可靠正確的科學證據，為經濟和社會帶來益處並促進永續發展。」於是，全球生物多樣性資訊機構（Global Biodiversity Information Facility, GBIF），這個由多國政府資助所成立的國際合作網絡及資料基礎建設在2001年正式成立。GBIF以「提供大眾無論何時何地，均能公開且自由存取有關地球上各種生物的資料」為宗旨，建立了整合全球的生物多樣性資訊開放平臺（GBIF.org）和相關資訊基礎建設。

臺灣生物多樣性資訊機構（Taiwan Biodiversity Information Facility, TaiBIF）於2004年在中央研究院生物多樣性研究中心成立及運作，是GBIF在臺灣的國家節點，主要任務為推動及建置臺灣生物多樣性資訊相關基礎建設，

包含生物多樣性資訊的開放分享、研究資料管理流程的建置與交流，以及資訊服務的設計與開發。TaiBIF不僅是推動國內生物多樣性資料整合及流通的領導者，更是臺灣與國際生物多樣性資訊連結的重要窗口。

在推動生物多樣性資訊整合中有兩個重要的部分，第一是關於資料的共享，透過與各生物多樣性相關主管單位及研究團隊討論資料管理的觀念和方法，提昇研究人員對於資料價值的認同，從而願意將資料的整理、備份、標準化，以及加強資料可讀與重複利用性等工作，納入日常研究活動當中。第二則是生物多樣性資訊服務及基礎建設的開發，透過制度的建立及軟體工程，確保資料的發布和取用流程中，均有合適的解決方案，讓不同使用者都可以找到並使用生物多樣性資訊。

推動資料開放與標準化的過程

無論是生態的保育或是研究，生物資料是一切的根基，所有的一切都必須先從能被看見、被找到開始，然而，開放生物多樣性資料其實不是一件容易的事情，要讓不同人、不同領域與研究的生物資料，能夠被分類、標準化，甚至是被國際看見，是需要很多不同層面的努力。首先，在學術研究的傳統認知上，為了預防自己努力的研究成果被別人覬覦，並不會隨意公開詳細的研究資料，尤其是原始資料。但當時間一久，未公開的資料因各種因素消

失的可能性就越高，像是硬碟損毀、僅留在某人的個人電腦、人員失聯等等。試想一下，這些研究資料的取得往往耗費了大量人力、物力、財力，然而在傳統學術論文中，這些資料通常經過轉譯、分析，原始資料沒有再次被使用的機會，最終被淹沒在個人電腦的資料汪洋中或消失不見。

再者，生物學家們研究各種不同的動物和植物，不論是研究目的、資料紀錄與田野調查的方式都不同，資料撰寫的格式也不一，因此很難將這些資料整理在一起進行分析比較，就像大家都在寫不同的語言，無法溝通一樣。

那麼，要怎麼能讓大家的資料，能夠跨領域的被不同人理解並使用呢？

生物多樣性資料的標準格式－DwC達爾文核心集

「FAIR原則」於2016年首次正式發表在《科學資料》（Scientific Data）上，成為一個被廣泛接受並在科學研究社群中推動的標準，以提升數據的互通和可利用性。FAIR原則是由四個互相關聯名詞組的首字縮寫，描述開放資料必須要符合找得到（Findable）、拿得到（Accessible）、可互用（Interoperable）、可再用（Reusable），這樣才能更容易應用在不同研究與政策制定上。

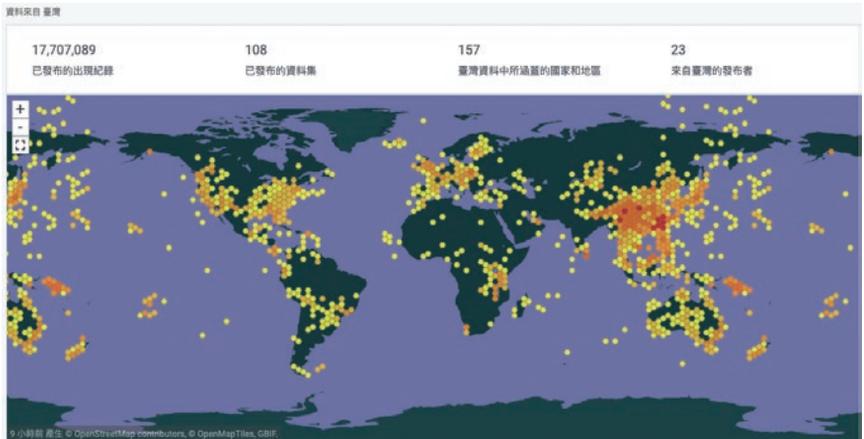
為了達成資料可互用的目的，讓不同研究者的生物資

料可以被整合使用，資料專家們，就生出了達爾文核心集（ Darwin Core, DwC；<https://www.tdwg.org/>），DwC是一個用於生物多樣性資料的標準格式。它的目的是讓科學家能夠共享、整理和比較資料，特別是物種的分類與分布資訊。

DwC就像是一種通用的語言，它定義了生物多樣性資料應該以怎樣的方式組織和呈現，提供一套共同的規範，使得不同的研究者可以使用相同的結構來描述他們的資料。這樣一來，無論資料來自哪個研究者或來源，我們都可以將它們整合到一個統一的資料庫中。標準化格式對於生物學研究的進展非常重要，因為它可以促進合作、重複實驗和跨領域的研究。透過DwC將生物多樣性資料標準化以後，不僅僅是提升了資料品質，更增加了資料的可用性，有助於我們對不同類型研究的認識與應用。

資料嘉年華

當開放資料量累積到一定程度，便能彌補許多研究資料在時間、空間、生物類群尺度上的不足，並藉由大數據分析、模式預測等將資料再利用，支持更多大尺度的科學研究。臺灣目前是亞洲生物多樣性開放資料第二多的國家，到2023年11月15日為止，在GBIF的資料平台上，我國已有23個資料發布單位開放了107個資料集，包含17,707,089筆物種的出現紀錄（圖一）。



圖一、GBIF資料平台上臺灣的發布狀況。（圖片來源：GBIF）

近年科技的迅速發展，迎來全球資料與資訊的爆炸和隨之而來的模式轉變，讓包括生物多樣性相關領域的科學研究及應用更趨向資料導向。伴隨這種轉變中的研究與應用模式也整合了許多新興的生物多樣性資料與資訊來源，以下為大家分享臺灣不同的主題案例，包含新興資料類型與生物資料的應用。

案例一：新興資料類型—eDNA帶你看遠古至今

eDNA（Environmental DNA）即環境DNA，是一項生物檢測的技術，用於從自然環境中採樣樣本中所含不同生物DNA片段，識別當地的生物種類，而不需直接觀察或捕捉這些生物。以往，研究人員為了蒐集在田野間物種的

資訊，往往需要耗費許多人力與時間成本，穿梭於不同環境，並採集樣本，然而隨著eDNA技術的出現，在採集與辨識時可以更加的快速，且透過小份的樣本就可獲得更多的物種資訊。透過提取樣本中的eDNA並進行分子生物技術的分析，有助於對當地生物存在的了解，甚至同時還可以用來研究古生物和已滅絕的物種。



圖二、於猴洞溪進行水質檢測與採樣。（拍攝者：王敏真）

一個由中央研究院生物多樣性國際研究生學程與臨海研究站合作的課程中，修課學生使用eDNA方法檢測河流生態系統中的生物物種，並比較從上游的猴洞坑瀑布（圖

二），經過猴洞溪、下埔排水線，到下游竹安出海口物種組成的差別。修課學生將採樣水中所定序出來的DNA序列比對基因資料庫，辨識出動物、植物、真菌、原生物與色藻等不同的物種，並在TaiBIF的協助下，將基因序列公開於歐洲核酸序列資料庫（ENA），並將生物物種的出現紀錄公開在GBIF平台[1]。由於cDNA技術仍在發展中，且目前的基因資料庫還不完整，因此不是所有的序列都有辦法成功辨識出來，但透過資料的開放與妥善的保存，這些暫時無法辨識物種的出現紀錄與序列，待未來基因資料庫更完整時，將可以重新分析比對，不會「浪費」這些寶貴的資料。

案例二：新興資料類型之二－解謎森林中的大合唱，用AI辨識鳥音

在AI正夯的時代，你打算如何使用這些工具來幫助你的研究呢？為了解氣候變遷對森林中鳥類的潛在影響，農業部生物多樣性研究所的吳世鴻副研究員與玉山國家公園合作，在國家公園內不同海拔高度的森林架設錄音機，長期收錄環境中的聲音，作為監測生物活動與分布的工具（圖三）。

錄音聽起來容易但其實不然，研究人員需要從大量且複雜的錄音檔中，擷取出監測目標物種的聲音及相關資訊，再進行後續的分析。面對挑戰，研究人員利用深度學

習技術開發了一種名為SILIC（生物音智慧辨識與標記系統）的人工智能工具，可以透過聲音識別物種，經過多次優化，目前已可識別169種臺灣本土野生動物，包括鳥類、青蛙、哺乳動物和爬行動物[2]。



圖三、在樹上安裝自動錄音機以紀錄環境中的聲景。（拍攝者：吳世鴻）

這項研究針對自動辨識正確性高達95%的7種山地森林常見鳥類，將所辨識出超過6百萬筆的出現紀錄透過TaiBIF開放於GBIF的資料平台，成為全球第一份由AI辨識聲音而產生並開放的物種出現紀錄資料集。「期望我們的

資料集可以填補高山森林中鳥類細緻時序活動模式的資料缺口，並有助於氣候變化對高山森林生態系統影響的研究做出貢獻。」研究人員說。透過發展處理及管理這些新興資料的方法與流程，將使開放資料政策和結構邁向創新並加速發展。

那麼，又如何將開放資料應用到研究上？

為了推動臺灣開放資料的跨領域研究，TaiBIF也與許多夥伴單位如林業及自然保育署、生物多樣性研究所共同合作，讓不同主題的研究能夠蓬勃發展，研究內容除了分類學、生態學及生物地理學等基礎研究之外，也有探討氣候變遷、外來入侵種衝擊等重要社會與環境議題的研究。截至2023年，依據GBIF的文獻追蹤系統，在全球使用GBIF資料的研究論文中，有146篇來自臺灣學者的參與，另有至少47篇使用了從臺灣發布的開放資料。在此分享幾篇應用TaiBIF和GBIF上開放資料的案例：

案例三：透過自然史典藏資料，探討物種分布模式

國立臺灣博物館研究助理吳士緯，在擔任中央研究院生物多樣性研究中心的博士後研究員時，與研究員沈聖峰等人，運用農業部生物多樣性研究所在GBIF所發布超過兩萬張的蛾類影像資料，結合AI工具尋找蛾類色彩多樣性隨著環境梯度變化的相關性[3]。

他們發現隨著分布海拔的下降，蛾類的身體顏色變

得更為多樣，且身體和前翅的明暗度在較高海拔會變得較低，反映出蛾類透過身體顏色的改變適應不同海拔的環境。對於預測稀有蛾類的分布和生物與環境之間的關係，尤其是對氣候變遷的反應有很大的幫助。

案例四：結合公民科學資料，驗證動物遷徙模式

海拔遷徙在鳥類中是一個普遍但重要的行為，對於生活在山區的物種來說，這樣的行為讓鳥類適應在環境中不同海拔高度上資源分布的季節性變化。然而過去由於觀測資料的缺乏，對影響這項行為的因素了解有限。

中央研究院生物多樣性研究中心的研究助理蔡佩妤和副研究員端木茂甯，與生物多樣性研究所的副研究員柯智仁等人合作，透過公民科學eBird的開放資料，研究分析臺灣繁殖鳥類在不同季節時海拔高度分布的變化。這研究讓我們能夠觀察到以往難以捉摸的鳥類遷徙模式，同時，也利用鳥類的行為和型態特徵來驗證多個關於海拔垂直遷徙的假設，找出影響此項行為的關鍵環境因子[4]。

案例五：運用氣候及物種開放資料，預測氣候變遷的影響

過去幾十年來，熊蜂的全球分布和數量大幅減少，主要威脅包括殺蟲劑接觸、棲息地喪失、病原體感染、入侵物種和氣候變遷。在這些因素中，氣候變遷對高海拔熊蜂

物種尤其不利，與低海拔地區更常見的物種相比，它們的生態幅度有限，而且負面影響更大。

生物多樣性研究所副研究員呂明倫及助理研究員黃靜宜，從GBIF及其他開放資料庫的資料進行研究分析，評估未來氣候變遷對臺灣特有種——信義熊蜂（*Bombus formosellus*）的負面影響，並找出在此衝擊下合適的山區棲息地。結果顯示臺灣山區的暖化越趨顯著，未來將不再適合信義熊蜂棲息，並可能因此讓信義熊蜂滅絕[5]。

現在進行式

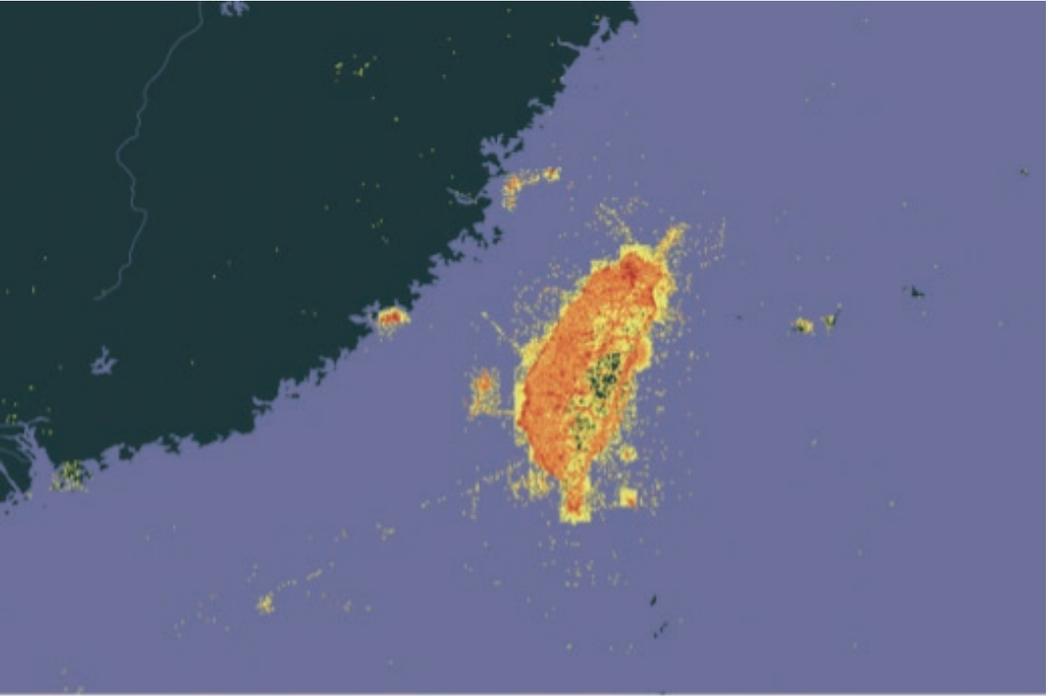
除研究外，國內政府部門也愈來愈重視開放資料的應用，並著手跟學者合作分析資料並作為保育決策的參考。案例如臺灣繁殖鳥類大調查，是針對臺灣有繁殖族群的鳥類所進行的公民科學計畫，每年號召全臺數百名鳥友定期定點進行鳥類調查。生物多樣性研究所與中華民國野鳥學會運用「臺灣繁殖鳥類大調查」的資料，分析2011～2019年間的100種繁殖鳥類在臺灣的數量變化趨勢，並建置了「臺灣森林鳥類指標」及「臺灣農地鳥類指標」兩項國家級指標，以此作為監測臺灣繁殖鳥類生存狀態及保育成效的重要儀表板，用來反映臺灣環境現況，也讓臺灣成為亞洲第二個發布複合物種指標的國家[6]。

2021年中央研究院生物多樣性研究中心與國內管理生物多樣性資料庫的保育、典藏機構及學術研究等政府單

位，成立了臺灣生物多樣性資訊聯盟（Taiwan Biodiversity Information Facility, TBIA），共同推動生物多樣性資料的流通整合，以支援研究與保育應用，並由TaiBIF協助建置跨單位的資料入口網（<https://tbiadata.tw/>），讓使用者可以從單一網站查詢並取得所有聯盟夥伴單位所有的開放資料。目前聯盟成員包括農業部的林業保育署、生物多樣性研究所和林業試驗所、海洋委員會海洋保育署、內政部國家公園署、國立臺灣博物館、經濟部水利署、國立自然科學博物館與中央研究院數位文化中心。隨著聯盟的持續擴大，將來預計納入更多類型的生物多樣性資料，增加更多應用的可能性，期許TBIA逐漸成為全國性的生物多樣性資料庫網絡，而TaiBIF團隊則在其中提供整合各單位資料所需的軟硬體技術及基礎建設，支援生物多樣性的應用發展。

隨著資訊基礎建設的完善和大量資料的累積後，我們也將面臨下一個挑戰，包括彌補資料在時空間與類群分布上的空缺、處理新興的生物多樣性資料與傳統資料的結合，提升資料品質並增強資料在研究和政策制定上的應用。

面對這些挑戰，TaiBIF不僅持續針對上述問題提出解方，也憑藉過去的經驗和技術，引領國內生物多樣性資訊學的發展、協助政府達成永續發展目標願景，並在GBIF亞洲區域事務中扮演重要角色，提升臺灣在國際生物多樣性資訊研究領域的影響力（圖四）。不論未來的研究應用如



圖四、每一個光點都代表了一筆GBIF收到的生物資料。在臺灣發布者的努力下，累積了足以點亮臺灣輪廓的巨量資料。（OpenStreetMap®, GBIF.org, CC-BY-SA 2.0）

何革新，TaiBIF將持續給予最大的支持，成為生物多樣性資訊基礎建設最強而有力的推動者。

參考資料

1. European Nucleotide Archive (2023, April). *eDNA along riverine zonation of Houdong river, Yilan, Taiwan*. <https://www.ebi.ac.uk/ena/browser/view/PRJEB60905>
2. Wu S.-H. et al. (2023) An acoustic detection dataset of birds (Aves) in montane forests using a deep learning approach.

- Biodiversity Data Journal*, 11, e97811.
3. Wu, S. et al. (2019). Artificial intelligence reveals environmental constraints on colour diversity in insects. *Nature Communication*, 10, 4554.
 4. Tsai, P. et al. (2021). New insights into the patterns and drivers of avian altitudinal migration from a growing crowdsourcing data source. *Ecography*, 44, 75-86.
 5. Lu, M. & Huang, J. (2023). Predicting negative Effects of Climate Change on Taiwan's endemic Bumblebee *Bombus formosellus*. *Journal of Insect Conservation*, 27, 193-203.
 6. Lin, D. et al. (2023). Taiwan's Breeding Bird Survey reveals very few declining species. *Ecological Indicators*, 146, 109839.

後記

生物多樣性研究中心的「臺灣生物多樣性資訊機構TaiBIF」在GBIF共享超過17,727,039筆物種出現紀錄，TaiBIF每年都會舉辦工作坊，手把手教學資料開放流程與資料清理，歡迎更多人加入，隨著資料的開放，也期待未來有更多不同層面的應用與研究。

作者簡介



吳佳奇

撰文時為中央研究院生物多樣性研究中心專題中心助理，同時也是臺灣生物多樣性資訊機構（TaiBIF）臉書粉絲頁小編，希望透過簡單易懂的方式讓開放資料觀念更加普及。



何芷蔚

臺灣生物多樣性資訊機構（TaiBIF）博士後研究學者。國立臺灣師範大學與中央研究院生物多樣性國際研究生學程博士（TIGP Biodiversity Program）。社團法人台灣海龜點點名協會共同創辦人。目前研究興趣主題圍繞在海龜與資料上：真菌感染海龜蛋、海龜水下覓食族群生態、海龜便便以及所有調查資料如何更有效率的做分析、格式化與開放。有點資料潔癖。



端木茂甯

中央研究院生物多樣性研究中心副研究員，代理系統分類與生物多樣性資訊專題中心執行長，專注於生物多樣性的時空動態與影響機制的研究。他是一位外表和善寡言但內在卻洋溢研究熱情的科學家。每天以一身橫條 polo 衫靜坐在電腦前的姿態，徜徉於資料的海洋，搜尋著隱藏的知識寶藏，沒有任何其它嗜好能夠分散他對科學的執著與追求。這份熱愛持續引領他探索未知的領域並綻放出獨特的光芒。