

# 生物多樣性資料品質的 評估機制與應用

文／邱祈榮（國立臺灣大學森林環境暨資源學系副教授）  
陳品元（通訊作者 | 加州大學戴維斯分校博士生）

資料品質 (Data Quality) 是指資料能滿足使用者需求的符合程度，包含資料適合使用或有潛力做後續使用的程度 (English, 1999; USGS, 2004)。資料品質管理準則 (Data Quality Principle) 已成為諸多領域的核心工作，包含商業、醫學、地理資訊系統 (GIS) 與遙測等 (Gad and Taulbee, 1996; Zhang and Goodchild, 2002; Lunetta and Lyon, 2004; SEC, 2017)。來自政府部門、博物館、植物園等的長期資料，能反應長時間一定範圍內的物種多樣性、豐富度、棲地與環境的變化，然而，此概念目前仍未被政府部門、博物館與公民科學界廣為接受 (Chapman and Busby, 1994)。

完善的品質管理機制能梳理上述機構龐雜的資料品質 (Chapman, 2005)。在進行流程前，可依照資料的取得方式、資料的性質、資料的呈現方式等方法進行初步分類 (林惠玲、陳正倉, 2009)，例如將資料區分為具有族群、物種、分布位置等資訊的生物特性資料，以及具有環境因子等資訊的非生物特性資料。待資料經過初步分類後，統計學上常藉由以下項目評估資料品質，包含準確性 (Accuracy)、精確性 (Precision)、不確定性 (Uncertainty) 等

(Chapman, 2005; Hill *et al.*, 2010; Askham *et al.*, 2013)。

現存的臺灣生物多樣性資訊機構 (Taiwan Biodiversity Information Facility, TaiBIF) 能負責整合臺灣生物多樣性的相關資訊 (中央研究院生物多樣性研究中心, 2001)；另外政府部門也有將公開的年報資料上傳至官方網站。臺灣目前雖已具備龐大的生物多樣性資料，但由於資料龐雜，尚有資料品質參差不齊、缺乏資料品質評估機制等問題，仍然缺乏反映全國生物多樣性現況的資料。因此本研究旨在整合國家層級的臺灣生物多樣性資料，同時建立資料篩選機制與品質評估流程，以瞭解臺灣生物多樣性資料的品質與適用範圍。

## 資料收集與品質評估機制的建立

### 評估資料品質的 3 大前提

臺灣生物多樣性資料相當多元，這些資料主要來自各政府主管機關，例如行政院農委會林務局、內政部營建署、特有生物保育研究中心 (下稱特生中心) 等，以及開放的資料平臺，例如 eBird、全球生物多樣性資訊機構 (Global Biodiversity Information

Facility, GBIF) 等。在獲得管理單位的同意與授權後，應依據下列 3 項前提評估所蒐集的資料是否足以進行資料品質檢核 (圖 1)：

- 確認是否有監測單位負責資料蒐集與彙整的工作
- 確認該單位所蒐集的資料是否為長期且連續的資料
- 確認此長期蒐集的資料所依據的監測方法是否具有的一致性且標準化

依據資料類型區分為生物資料特性與非生物資料特性 2 種類別。其中，非生物資料特性是指環境資訊、經營管理、財務資訊等；生物資料特性則是包含物種、族群、棲地、空間及時間，共 5 項資訊 (圖 1)：

- 物種資訊是指物種的鑑定
- 族群資訊為族群類別與調查數量
- 棲地資訊是指物種的分布狀態
- 空間資訊為樣區尺寸與調查取樣範圍
- 時間資訊為資料蒐集頻率

### 資料品質檢核流程的建立

為客觀與專業地檢核資料品質，研究人員會優先瞭解資料的特性與細部產製流程，將符合上述 3 面向評核的長期監測資料，

此外，過去許多研究已指出資料的精確度、準確性、完整性與一致性等，均為有效評估資料的品質的重要元素 (Chapman, 2005; Hill *et al.*, 2010; Askham

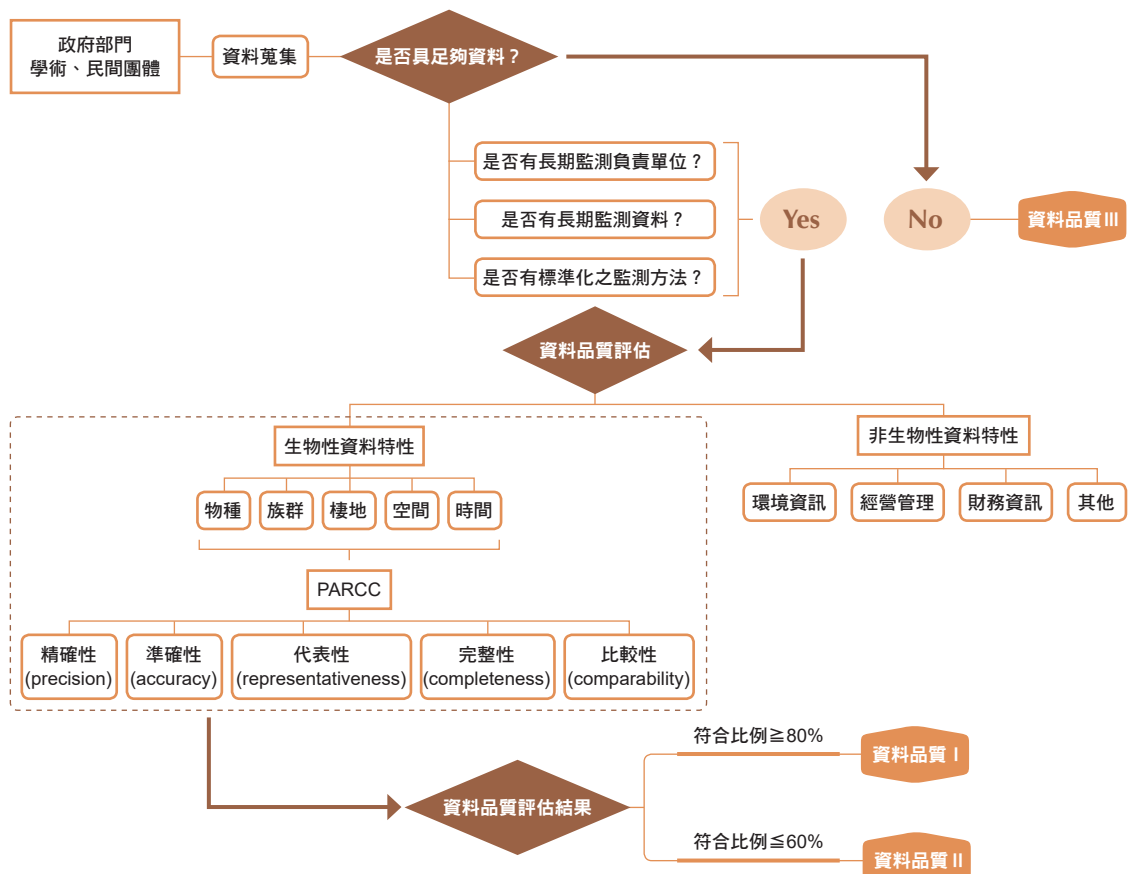


圖 1 資料品質評估前提與檢核流程

et al., 2013)。因此完成資料特性的區分後，相關計畫的負責人與研究助理會使用 PARCC 的架構，針對上述 5 項生物資訊的精確性 (Precision)、準確性 (Accuracy)、代表性 (Representativeness)、完整性 (Completeness)、比較性 (Comparability) 進行評估 (圖 1)：

- 精確性描述資料的變異度大小，意味著同一參數重複測量的一致性，一般使用的統計值為標準誤差值，其數值越小越好 (Chapman, 2005; Hill et al., 2010)。
- 準確性為是指資料組具有的偏差大小，意味著實際值與估計值的差距，一般使用的統計值為相對誤差值 (Chapman, 2005; Askham et al., 2013)。
- 代表性為取樣調查的樣本資料是否能準確地反映出族群母體特性，常見的代表性考量有取樣母體、時間與空間分布的調查性 (Askham et al., 2013)。

- 完整性指成功調查到的有效數據與原本規劃調查樣本數量的比率。有效數據與規劃調查之樣本間的差異越小越好，原先規劃卻未能調查到的樣本會影響精確度與準確度，降低資料的可信度 (Askham et al., 2013)。
- 比較性主要反映資料的一致性，包括長期資料調查所反映對象、調查方法、分類方法與資料記錄單位。若針對相同調查或監測項目，可能因分類系統、記錄方式或單位不同而無法比較 (Chapman, 2005; Askham et al., 2013)。

再來會依據五乘五的表格評估每項生物性資料的 PARCC 結果，並計算符合占全部評分項目的比例，得出百分比呈現評估結果 (圖 1)。關於各項生物性資料的符合標準詳見表 1。最後，完成 PARCC 的評量後，會將資料分為 3 大等級：等級 I 的資料品質尚可且能穩定供應，等級 II 的資料

表 1 生物資料特性與 PARCC 評量的符合標準

資料特性 PARCC	符合標準				
	物種	族群	棲地	空間	時間
P (精確)	調查者修習專業課程或受相關培訓機制	調查者修習專業課程或受相關培訓機制	調查者修習專業課程或受相關培訓機制	使用一致座標系統及空間精度設定	明確規定調查時間與天氣狀況
A (準確)	具複查機制，調查者會收集照片 / 標本	具複查機制，調查者會收集照片 / 標本	具複查機制，調查者會收集棲地照片	調查者使用合適器材記錄座標點位	調查者使用合適器材記錄時間
R (代表)	調查者遵循實驗設計的調查方法	調查經實驗設計	實驗設計明確規定棲地類型	樣區選擇與尺度經實驗設計	調查時間與頻率經實驗設計
C (完整)	完成所有預選樣區的物種調查	完成所有預選樣區的族群調查	預選樣區涵蓋所有棲地類型	完成所有預選樣區的調查	能定期執行調查與更新資料
C (比較)	調查紀錄表使用一致物種分類系統	調查紀錄表使用一致調查方法	調查紀錄表使用一致棲地分類系統	調查紀錄表採用統一座標系統	調查紀錄表需記錄日期與時間

品質欠佳且無法定期供應，等級III因不具足夠資料，無法進行 PARCC 評估（圖 1）。

## 資料品質評估結果與應用

### 資料蒐集來源

本研究蒐集的臺灣生物多樣性資料主要來自政府部門的公開網站、學術及民間團體以及公開資料平臺，包含行政院農委會林務局、行政院農業委員會動植物防疫檢疫局、行政院環境保護署、內政部國土測繪中心、內政部營建署、內政部營建署城鄉發展分署海岸復育課、經濟部水利署、特生中心、東華大學自然資源與環境學系兩棲保育研究室、中華民國野鳥學會、黑面琵鷺保育學會、eBird 等，共蒐集 30 項資料。

### 資料品質評估與分級結果

初步評估結果顯示，有 9 項資料因為缺乏長期監測單位、監測資料與標準化監測方法，因此，無法進行資料品質評估。本報告將針對資料涵蓋 4—5 種生物訊息的資料—常見繁殖鳥類、小花蔓澤蘭，進行詳細的 PARCC 評估結果說明。

#### ■ 鳥類

是一般民眾熟悉的生物類群，具備長期監測資料且被認為是絕佳的環境健康反映指標（Padoa-Schioppa *et al.*, 2006; Goodale *et al.*, 2008; Berglund, 2018），因此常見繁殖鳥類指標的發展也格外重要。繁殖鳥類大調查（Breeding Bird Survey, BBS）結合政府機關與民間組織，透過公民科學家參與，以相同規範的調查方法監測臺灣常見鳥類的族群數量。資料品質評估結果顯示 25 項評估項目中，21 項評估結果符合，4 項評估結果則為不符合，符合比率為 84%（表 2）。

#### ■ 小花蔓澤蘭

是一般民眾熟知的外來植物，其環境適應性強且生長速度快，會阻礙、纏擾其他植物生長進而對棲地環境造成危害。小花蔓澤蘭覆蓋面積的監測，是由特生中心函請各縣市政府的相關單位填報相關資料，再由林務局編製「臺閩地區外來植物物種覆蓋率—小花蔓澤蘭」統計表。該資料缺乏空間資訊，評估結果顯示 20 項評估項目中，僅 5 項符合，符合比率為 25%（表 3）。目前所收集 30 項



■ PARCC 評估結果以常見繁殖鳥類（圖為烏頭翁）及小花蔓澤蘭進行示範並製成表格說明。（左／許宸提供）（右／林務局提供）

表 2 常見繁殖鳥類的資料品質評估狀況

資料特性	物種	族群	棲地	空間	時間
P 精確	調查者受專業培訓且具雙筒望遠鏡、紀錄紙、筆、相機與圖鑑	調查者受專業培訓且具雙筒望遠鏡、紀錄紙、筆、相機與圖鑑	調查者受專業培訓能精確判定主/次要棲地類型	使用空間精度設定相同的經緯度座標系統	規定於日出後，在非強風/大雨/濃霧的天候調查
評估	符合	符合	符合	符合	符合
A 準確	具複查機制且調查者會收集照片	具複查機制且調查者會收集照片	具複查機制且調查者會拍照記錄	使用車用/手掌型 GPS 記錄座標	調查者正確紀錄調查時間與日期
評估	符合	符合	符合	符合	符合
R 代表	工作手冊經實驗設計	工作手冊經實驗設計	工作手冊包含約 35 種棲地類型代號	樣區選擇與尺度經實驗設計	規定在日出後調查，各取樣點有 2 次調查，間隔至少 2 周
評估	符合	符合	符合	符合	符合
C 完整	過去規劃約 600 個預選樣區，已累計 282 個樣區	過去規劃約 600 個預選樣區，已累計 282 個樣區	過去規劃約 600 個預選樣區，已累計 282 個樣區	過去規劃約 600 個預選樣區，已累計 282 個樣區	定期執行調查與更新資料
評估	不符合	不符合	不符合	不符合	符合
C 比較	調查紀錄表使用一致鳥種分類系統	調查紀錄表使用一致調查方式	調查紀錄表使用一致棲地分類系統	皆使用經緯度座標系統	需在調查紀錄表標明日期與時間
評估	符合	符合	符合	符合	符合

表 3 小花蔓澤蘭的資料品質評估狀況

資料特性	物種	族群	棲地	時間
P 精確	調查無委託專業人士，也無規定調查器材與裝備	調查無委託專業人士，也無規定調查器材與裝備	記錄土地別代號及文字描述地點	無明確規定調查時段與天氣狀況
評估	不符合	不符合	符合	不符合
A 準確	無資料複查機制，未收集照片	無資料複查機制，未收集照片	無資料複查機制，未收集照片	調查者正確記錄調查時間與日期
評估	不符合	不符合	不符合	符合
R 代表	調查方法未經實驗設計	調查方法未經實驗設計	調查方法未經實驗設計	調查時間未經實驗設計
評估	不符合	不符合	不符合	不符合
C 完整	無法得知有效鑑定物種比率	僅有林務局資料，未整合其餘轄區資料	僅有林務局資料，未整合其餘轄區資料	無記錄調查頻率
評估	不符合	不符合	不符合	不符合
C 比較	調查對象為單一植物物種	缺乏一致調查方法	記錄土地別代號及文字描述地點	調查表格記錄年/月/日
評估	符合	不符合	符合	符合

表 4 資料品質分級結果

資料品質等級	資料品質特性	資料數(項)
I	<ul style="list-style-type: none"><li>資料提供穩定</li><li>時間資料的完整性高</li><li>PARCC 的符合比例<math>\geq 80\%</math></li></ul>	15
II	<ul style="list-style-type: none"><li>無法穩定提供</li><li>時間資料的完整性較低</li><li>PARCC 的符合比例<math>\leq 60\%</math></li></ul>	6
III	<ul style="list-style-type: none"><li>缺乏相關方法學</li><li>未建立資料收集機制</li><li>無法進行 PARCC 評估</li></ul>	9

資料中，屬於等級 I 者有 15 項、等級 II 者有 6 項、等級 III 者有 9 項（表 4）。

由資料品質評估結果可知，時間、空間與族群資料的代表性、完整性是目前臺灣生物多樣性資料面臨的問題。

### 資料品質分級的應用

將資料進行評估與分級，能簡單且明確呈現各項資料的品質與特性。若欲針對相關資料進行分析，即可透過此機制篩選符合需求的資料，例如本研究欲針對臺灣生物多樣性的長期趨勢變化分析，然而該分析需探討導致趨勢變化的因子，因此本研究選擇能穩定更新且品質尚可的等級 I 資料作為分析對象。另外，研究人員亦可針對注重的內容，針對 PARCC 的各個項目進行加權，以篩選適合的資料。

除了針對資料品質進行篩選外，本研究採用的資料品質評估流程可以幫助研究者與分析人員瞭解資料包含哪些生物性訊息。由評估結果可知，並非所有資料均涵蓋物種、族群、棲地、空間與時間等訊息，部分資料僅包含 2 至 3 項資訊。因此在資

料品質評估流程中，透過先將資料特性做初步分類的方式，確認各項資料的設置目的與所含資訊，藉此提升研究者在進行資料分析前置作業時的便利性與工作效率。

### 實際運用面臨的課題

然而即使已進行初步篩選，也完成資料品質評估，取得與呈現的資料仍存在無法反應實際生物多樣性現況的問題。研究團隊在探討影響天然河岸長度的趨勢因子時發現，權責單位提供的資料母體並非臺灣總河岸長度，而是僅涵蓋下游主河道，因此即便是在經過初步品質評估結果尚可，仍有可能出現無法反應實際狀況的情形。為避免上述狀況發生，瞭解資料的產製流程、確認來源、分析與收集方式扮演相當重要的角色。建議透過親自或電話訪視資料提供單位，明白其細部產製流程，包含資料的調查方法、收集方式、分析前置作業、分析方法等，如此有助於釐清資料特性與適用範圍（邱祈榮等，2019）。

總體而言，在長期生物多樣性資料的收集與彙整方面，資料品質管理無疑扮演極為重要的角色。資料品質評估結果指出 PARCC 的資料評估流程能將資料品質進行分級，但此流程仍需配合對於細部產製流程的瞭解，才能更精確解讀與評估資料品質。此外，資料盤點結果顯示臺灣已有生物多樣性資料足以進行資料評估，若欲提升未來的資料品質，短期內建議可透過明確記錄資料收集方法、調查時間與定期更新資料的方式，來優化時間性的資料品質。🌱

（參考文獻請逕洽作者）