生物多樣性資料流通課程

GBIF 秘書處

版本 12, May 2021

目錄

課程說明

這門課程中，將讓參與者學習到：如何依照共通的操作標準來進行作業，藉以更有效率地去規劃、發布與應用生物多樣性資料。 這門課程由 European Union 所贊助的 Biodiversity Information for Development (BID) 計畫提出，希望透過這門課程，可以增進 GBIF 網路上的資料總量、豐富程度、與品質。

主題包含：

- 專案管理
- 資料擷取
- 資料管理
- 資料發布

這門課程包括影片的操作說明、隨堂小考及範例練習。當這門課以實體、或線上視訊工作坊進行時，我們鼓勵您在其中多多進行團隊互動。

讀者

這門課程是為 生物多樣性領域相關的研究者、技術員、或政策單位 所設計，若您正在規劃、或具有想將自己單位的生物多樣性資料流通的需求，其中的操作說明將對您特別有用。

先決條件

1. Introduction to GBIF course
2. 此外，為確保課程的品質、與收穫，參與者最好擁有以下的先備知識、技能：
    - 了解如何使用基礎的電腦與網路操作、以及試算表的使用。
    - 對於地理與生物多樣性資訊有基礎的認識，如：了解地理和地圖之概念、知道基礎的分類學與命名規則。
    - 願意將相關訓練教材根據當下所處之脈絡進行修改，並同時不改其教學價值，藉此將工作坊學習到的知識推廣給計劃中的工作夥伴、協作者。
    - 良好的英文能力。在此雖然我們努力提供其他語言的材料，但相關說明、影片仍會是英文。

學習目標

- 學習生物多樣性資訊的基礎、與關鍵概念(特別是生物多樣性數位資料的管理)
- 了解達爾文核心標準(Darwin Core Standard)的架構、與其組成內容
- 了解規劃生物多樣性資料數位化專案的不同階段，並學習如何據此應用到實際專案之上
- 審視資料流通策略，來評估潛在的差距、效率低下之處及遇到的陷阱。
- 能為特定機構規劃專屬的資料流通策略
- 學習分辨資料型態，及如何最有效地透過現有的軟體、工具、技術來獲取相關資訊
- 練習使用針對生物多樣性資料獲取而設計的軟體工具，並從中建立生物多樣性資料
- 建立資料品質的概念，並初步了解資料標準化、驗證、清理的相關工具介紹
- 學習以軟體工具評估一生物多樣性資料集的易用性
- 練習使用相關的資料清理工具

- 了解如何使用GBIF 資料整合發布工具(GBIF's Integrated Publishing Toolkit, GBIF IPT)來將生物多樣性資料公開發布
- 定義欲發布之生物多樣性資料中的資料主類型、副類型
- 選擇適當的延伸資料集(extensions)形式，並在GBIF 資料整合發布工具上發表資料集
- 協助他人規劃、獲取、管理及發布生物多樣性資料

認證

當您完成課程並通過認證作業測驗後，便有機會獲得官方的相關認證。認證將以 Open Badge 呈現，詳情請見Assessment and Certification

檔案下載

這堂課相關的下載檔案都在這個頁面上。 這些檔案將貫穿整門課程，並且也都具有單獨的頁面連結，因此您也可以選擇遇到時再播放、或下載。 其中，影片都已嵌入在課程頁面中，並也可以從YouTube播放(在YouTube上的影片多半具有字幕)。但若您無法播放、存取課程中嵌入的影片，建議您將這些.mp4檔案下載至電腦中播放。

影片

這些影片中將採用英文，且下載的影片中並不包含字幕。

基礎(一) Foundations1.zip (73.7 MB)

基礎(二) Foundations2.zip (90.2 MB)

規劃 Planning.zip (51.3 MB)

資料擷取 Capture.zip (63.1 MB)

資料管理 Management.zip (30.2 MB)

資料集發布 Publishing.zip (77.9 MB)

附錄：發表資料集論文 Appendix-Data-Papers.zip (97.5 MB)

練習資料

這個連結 compressed file (ZIP 37.7 MB) 中，包含了所有練習用的資料

練習試算表

這個連結 compressed file (ZIP 1.4 MB) 包含了所有的課後練習試算表。這個練習檔以英文寫成，且須以英文完成練習。

基礎

為確保每位參與者在接下來的資料流通課程前都至少有一定程度的先備知識，在這個環節中將進行基礎介紹說明。 其中包含：(1)使用的語言、專有名詞、部分基礎概念的定義，以及接下來課程中將使用到的功能與流程。(2)資料品質的概念與文件結構化的重要性 (3)安裝OpenRefine，在這一環節中會用上

專有名詞

定義

在這影片 (12:02)中，將介紹所使用的到的專有名詞。若您無法觀看嵌入在課程頁面上的影片，可以點此下載 download 並在電腦上觀賞。 (MP4 - 38.5 MB)

▶ https://www.youtube.com/watch?v=FZAF5Sy8Nsc *(YouTube video)*

軟體

在這影片
(05:58)中，將介紹生物多樣性資料流通過程中不同的環節，以及其對應功能的軟體。若您無法觀看嵌入在課程頁面上的影片，可以點此下載 download 並在電腦上觀賞。 (MP4 - 18.9 MB)

▶ https://www.youtube.com/watch?v=vYfDIgBBKXY *(YouTube video)*

資料結構

In this video (13:10), you will review the field and data types that hold data, the structures that help to organize and protect that data and what these mean for the integrity and security of your data.
在這影片中，將了解資料保存欄位、資料型態與資料結構，並進一步了解其對於資料保存與規劃上的助益，以及之於資料完整性與安全性的意義。若您無法觀看嵌入在課程頁面上的影片，可以點此下載 download 並在電腦上觀賞。(MP4 - 38.8 MB)

▶ https://www.youtube.com/watch?v=msnVbZvIy2E *(YouTube video)*

資料品質

在這影片 (12:26)中，將介紹所使用的到的專有名詞。若您無法觀看嵌入在課程頁面上的影片，可以點此下載 download 並在電腦上觀賞。 (MP4 - 44.5 MB)

▶ https://www.youtube.com/watch?v=5o7TcS2K7Cw *(YouTube video)*

以下為選讀文章——來自Arthur Chapman's guide 的 "Principles of data quality" (資料品質原理) 此為全文 Full document ， 參考資料及翻譯也可在GBIF.org上找到。

在近一步探討資料品質和它在物種出現紀錄上的應用前，尚有一些概念需要釐清，如：資料品質的意義、詞語上常被誤用的「正確性」與「精準性」間之定義與區別，與初級物種資料和物種出現紀錄資料的意義。

物種出現紀**錄**資料

物種出現紀錄資料(Species-occurrence data)包含：博物館與植物標本館中附在標本(或批次)上的標籤資料、一般的觀測資料、和環境調查資料。在這個範疇中，雖含有線段資料(環境調查中的

穿越線調查資料，如：沿河流之蒐集)、多角形資料(於特定區域內的觀測，如：國家公園)、網格資料(於規則網格中的觀測、調查記錄)等不同樣態的資料，但廣義上它們仍算作所謂的"基於點的"(point-based)資料。就此而言，廣義上我們討論的是具有地理參照資訊之資料，其中的地理參照資訊便能將之與特定時空位置連結起來——其中的地理資訊可以含有確切的地理座標(若有則如：經緯度、UTM)，或具對地區的具體描述(如：海拔、深度等)，並這些資訊都需搭配上時間之資料(日期、時刻)。

物種出現紀錄資料一般都會帶有詳細的分類名稱，但亦可能包含有尚未鑑別者。因此，「物種出現紀錄資料」一詞有時能和「初級物種資料(“primary species data”)」相互替換。

初級物種資料

「初級物種資料」(“Primary species data”)用來指最原始的標本蒐集資料，它包含了沒有空間屬性的命名分類描述資訊，如：名稱、系統分類、與不具有地理空間屬性的分類概念。

正確性與精準度

正確性(Accuracy)與精準度(Precision)經常被搞混，其因在於人們普遍不理解其中的差異。

「正確性」指的是量測值(或觀測值、估計值)與真值(或實際情況、或被接受為真的值，如：測量控制點的座標)的靠近程度。

「精準度」(又稱解析度)則可被分為兩種類型：　首先，統計精度指的是重複觀測結果與它們自身的接近程度，其與真值的接近程度無關，因此有可能有高正確性、但低精準度的狀況。　其次，數值精度指的是記錄有效位數的個數，並隨著電腦的出現更加明顯。舉例而言，資料庫中可以輸出經度/緯度至小數點下10位的準確度(約為現實中之0.1mm)，然而該記錄的解析度實際上並不優於10-100m(小數點下3-4位)，這常導致大眾對解析度與準確度產生了混淆、錯誤的認識。

除了帶有空間屬性的資料外，「正確性」與「精準度」這兩個詞也能被用在不具空間屬性的資料上。舉例而言，一個標本可能雖被鑑定至亞種的層級(高的精準度)卻分在錯的分類底下(低正確性)，或是僅鑑定至科的層級(高正確性、低精準度)。

資料品質

「資料品質」一詞具有多個面向，並同時與資料管理、建模分析、品質控管確認、儲存與呈現有關。在Chrisman (1991)及Strong et al. (1997)中分別指出，資料品質僅和資料的利用相關，並且無法獨立於使用者外評估。在資料庫中，一筆數據並沒有所謂「實際的」品質與價值(Dalcin 2004)，它們的潛在價值只有在某位使用者的利用下才進而實現——故也由此可知，資料的品質便與滿足使用者需要的能力息息相關(English 1999)。

Redman (2001)認為：一適合使用的資料必須是能被取得(accessible)、正確(accurate)、即時更新(timely)、完整(complete)、與其他來源一致(consistent with other sources)、與主題相關(relevant)、全面(comprehensive)的，並提供使用者適當程度的細節、且容易閱讀理解。

身為資料保管者，可以去思考該做什麼來面向更廣的使用者、及拓展資料庫的用途(如：增加其潛在的使用與相關性)。這類工作譬如：將資料欄位原子化、或新增地理參考資訊……等等——也是故，這樣的思考也勢必得在 使用性提升 vs. 投入的工作成本 中做出權衡。

品質保證/品質控管

一直以來，品質控管(QC)、品質保證(QA)之間的差異並不清楚。Taulbee (1996)認為強調若要達成品質目標，兩者缺一不可，並對此做出區別。她將品質控管(QC)定義為：根據內部的標準、流程、或程序進行的品質判斷，藉此來控制及監測品質；而品質保證(QA)則依據外部標準進行，藉由回顧過程中的種種行為、品質控管(QC)流程，來確保最終產品符合預定的品質標準。

在更以商業導向的方法中，Redman (2001)將品質保證(QA)定義成「以盡可能低的成本來滿足最重要客戶的需求，並為此設計來生產無瑕疵的資訊產品的所有行為活動」。

上述提到的術語如何實際應用尚不清楚，且在大多數的情境中，它們常被同義地使用，並用以描述資料品質管理的整體操作。

不確定性

「不確定性」(Uncertainty)可以視為是對「對一未知量的知識及資訊的不完整性」的衡量，若當完美的測量裝置存在時，則可確定它的實際程度 (Cullen and Frey 1999)。不確定性是觀察者對資料理解的性質，此性質與觀察者較有關，而非每份資料本身。在資料中永遠存在不確定性，若想透過記錄、理解、和視覺化這樣的不確定性來讓他人能理解並不容易。但總之，不確定性是用來理解風險及其評估的關鍵術語。

誤差

「誤差」(Error)一詞同時包含了資料的不精準(imprecision)與不正確(inaccuracy)兩個層面。有許多因素會造成誤差，但一般來說，誤差會被視為是隨機、或系統性的——隨機誤差指的是隨機抽樣造成測量值與真實值之間產生的偏差；而系統性誤差則來自值上的整體偏移，若從製圖學的視角看仍保有「相對的正確性」(Chrisman 1991)，是而在決定「適用性」時，在某些應用上系統誤差在可能是可被接受的。

舉個情境而言，是在使用不同的大地測量基準的時候。若在整個分析中只使用同個大地測量基準，可能不會發生大問題；但若在分析中使用了來自不同大地

測量基準、且有不同偏差的資料，(如：使用不同大地測量基準的資料集、或鑑定時參照了早期版本的命名法規)，則可能就會發生問題。

(Chrisman
1991)提到：由於無法避免誤差，因此應將它視為資料的基礎維度。只有當誤差與資料的同時呈現時，才可能去了解資料的限制、甚至理解現有知識的極限在哪。承上所述，需要計算整理空間、屬性、時間等三個維度的已知誤差，藉此更進一步了解資料性質。

資料驗證與清理

資料驗證(Validation)是檢查資料是否正確、完整、合理的過程。其中包含：格式檢查、完整度檢查、合理性檢查、限制檢查，資料回顧(藉此判別離群值及其他誤差。如：地理上、統計上、時間上、環境性的……等等)，以及相關領域專家(如：分類學家)對資料的評估。此外，這階段也會去檢查資料是否遵守適用的標準、規則和慣例。這些檢查常會將可疑紀錄標記、整理出來，並進行後續的確認。找出造成誤差的主因，並盡可能避免這些錯誤再度發生，是驗證與清理中的關鍵階段(Redman 2001)。

資料清理(Data cleaning)是修補上階段找出的錯誤的過程，其可同義於「資料清潔」(注意：有些人使用「資料清潔」一詞來同時概括資料驗證、清理兩個階段，但此處並非此用法)。這階段中，要避免無意間的資料損失、同時在更改資料時要特別小心。較好的做法是新舊並存(保留原始資料、及更正後的版本)，且並排保存在資料庫中，藉此若在清理時出錯時還能恢復成原始資料。

文件化

在這影片中(09:47)將介紹文件化(documentation)的重要性，及其與資料管理、發布的關聯。你將會學到資料映射(data mapping)、資料間相互關係、與後設資料(metadata)的概念。若您無法觀看嵌入在課程頁面上的影片，可以點此下載 download 並在電腦上觀賞。 (MP4 - 29.2 MB)

▶ https://www.youtube.com/watch?v=Z5-SYImGRGc *(YouTube video)*

資料數位化流程

在這部介紹資料數位化流程的影片中(07:20)，定義了以數位影像來將自然歷史館藏物件數位化的 五個階段，而這幾個階段也能較容易地應用在其他的生物多樣性資料來源上。若您無法觀看嵌入在課程頁面上的影片，可以點此下載 download 並在電腦上觀賞。 (MP4 - 26.8 MB)

▶ https://vimeo.com/120369455 *(Vimeo video)*

正如同影片中強調，數位化協議(digitization protocols)會隨組織而異，故確保採用的是被同意、記錄、且受到認可的協議是至關重要的。

在工作坊中，我們並不會教資料數位化本身，因為它很容易就變成長達一周的課程內容；相反地，我們會專

注在關於生物多樣性資料獲取的基礎介紹上。雖說如此，由於我們知道仍有許多夥伴對此感興趣，所以仍想提供給一些關於資料數位化的相關學習資源給你們。

組織資料數位化的方法有很多種，一開始可能讓人感到不知從何下手。記住重要的一點是，在大多數他人的數位化案例中，很可能就有著和你們打算數位化的同類型標本、物件。在這個練習中，我們會介紹一些實際的資料數位化流程資源，藉此幫助你們上路，這些練習也會成為未來工作坊中選擇、修改、評估工作流程的基礎。

這過程中的步驟可能包含：

- 數位化前預先管理、分層：這包含數位化前的資料準備工作，諸如：將資料指定獨一無二的辨識碼，這將幫助資料集引用時不會出錯，並讓所有衍生資訊保存在一起。
- 影像擷取：這包含了一定程度的事先規畫，除了影像擷取本身外(如：工作程序的定義、硬體的選擇)，亦包括影像如何在何處被儲存與處理。
- 影像加工：包含品質控制、版本轉換...等等。
- 數位資料獲取：資料數位化程序的核心，包含如何獲取資料庫中的關鍵資訊。影片中強調，雖然鍵盤最常見的輸入資訊方法，但越來越多的機構正轉向使用進階的資料輸入科技。
- 進行地理參考：地理資訊對生物多樣性分析非常重要，因此資料數位化專案都應盡可能地去萃取出更加正確的理資訊。

整合數位化生物館藏(Integrated Digitized Biocollections, iDigBio) (iDigBio)是美國聯邦自然資源局下生物多樣性館藏進階數位化(United States National Resource for Advancing Digitization of Biodiversity Collections) (ADBC) 的協調中心單位。iDigBio它們領導著全國範圍的工作規模，藉以讓數百萬計的生物樣本能以標準的數位格式提供給科學研究社群、政府單位、學生、教育者、及社會大眾。他們同時也有製作幾支討論資料數位化流程的影片。

若你想更了解不同樣本類型的特定工作流程，你可能也會對iDigBio系列的其他影片感興趣：

- “Digitizing Wet Collections” (4:34 mins) https://vimeo.com/120369690
- “Imaging Workflows for the Digitization of Dry-preserved Vertebrate Specimens” (7:25 mins) https://vimeo.com/160615629
- “Digitizing Herbarium Specimens” (7:34 mins) https://vimeo.com/120369768

軟體工具

回顧生物多樣性資訊使用之軟體工具

在課程活動中，我們將示範如何使用不同軟體工具(如：與資料數位化、資料品質或轉換相關)來工作。這些軟體或許你也曾經在日常工作中使用過。

為此，社群訓練者、指導者、與先前的課程參加者已先整理出一份清單，其中包含了生物多樣性資訊軟體工具的相關資訊，如：主網站連結、關鍵事實、及優缺點的整理。

Download Software-database-EN.xlsx. (23 KB)

當分析你不曾使用過的生物多樣性資訊軟體時，你需要根據你的目的來選擇、使用它。下述清單中列出了不同的評估面向，以供評估開始時的參考。這些面向是受到GBIF手冊 “Initiating a Digitisation Project” 中「characteristics of a good database solution」一章啟發而列出的，如以下：

- 價格：最重要的決定因素之一。除了軟體授權費用外，還需考慮運行的硬體、維護、升級、和使用時的專業知識。
- 功能性：你得先對「要軟體能達成什麼」的期待十分清楚，藉以確保它能有效地達成目標。不要分心在附加功能上，它們可能會讓軟體不必要地變得更加複雜。

- **穩**定性：有些解方已經在市面上好一陣子，且被實體的機構、公司支援，它們更可能是沒有bug、或擁有良好的系統來解決任何出現的問題。它們也更有可能更新並移植到更現代的作業系統上。
- 可**擴**展性：
  有些軟體在剛開箱展示時表現良好，但一段時間後、或使用大量資料時、或同時多人存取時的表現卻會下降。請參考網上其他使用者的意見。
- 整合性：請確保軟體能讀取你使用的資料格式，且也能生產出你需要的格式。資料轉換可是件耗時的工作。
- 語言支援：讓每個軟體使用者都能理解它的介面、且也都了解可能會使用到的檔案，這點十分重要。
- 文**檔**與技術支援： 確保探索完既有的文檔與技術支援機制，你將確定在某刻你會需要它們。
- 學習曲線：有些軟體可能需要特別的訓練來學習如何使用，但有些則更加直覺性、且能在使用中學會操作(透過內建的幫助系統)。

安裝 OpenRefine

安裝後續課程活動所需之軟體工具

OpenRefine
是個專門用來改善資料集整體品質的工具，其中含有許多能處理表格資料的功能。這個應用程式能在你的電腦上做為小型伺服器運行，且為了使用它，須將你的網頁瀏覽器連上該伺服器的位址。因此，請將OpenRefine想做為一個個人、私人的網路應用程式。

在資料數位化課程的一部份(尤其在實作單元)中，將使用OpenRefine 進行，因此需要會先在你的桌電中安裝完畢。若你是位熟練的電腦使用者，你可以依循下面的這些步驟來在電腦上安裝軟體；但若你對此並沒有信心，請尋求協助。更多細節請參考 OpenRefine download page 。

在安裝時，可能會需要用到系統管理者密碼。

安裝需求

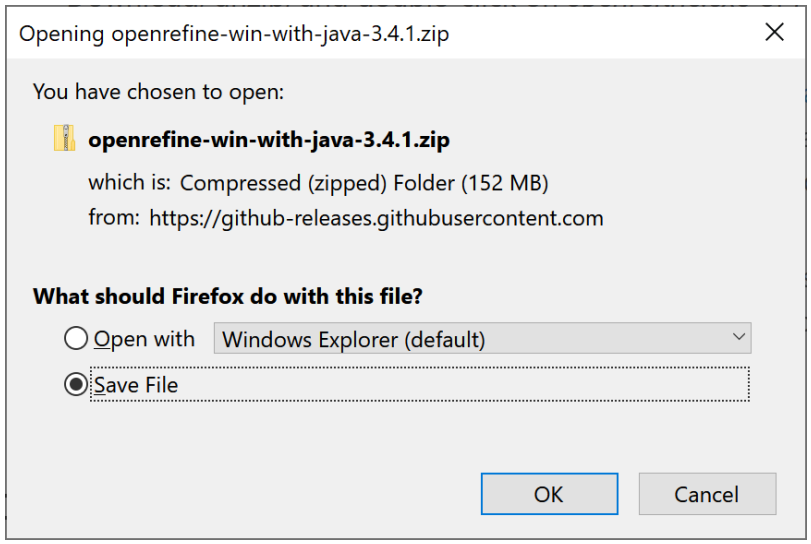
1. Linux 使用者： 需已安裝Java JRE
2. 已安裝如Google Chrome、 Microsoft Edge 或Mozilla Firefox 等網頁瀏覽器。請注意，它並不支援Internet Explorer。

最新的OpenRefine穩定版本為發布於2020/9/24的3.4.1版。更多安裝說明請見 https://docs.openrefine.org/manual/installing/ 。

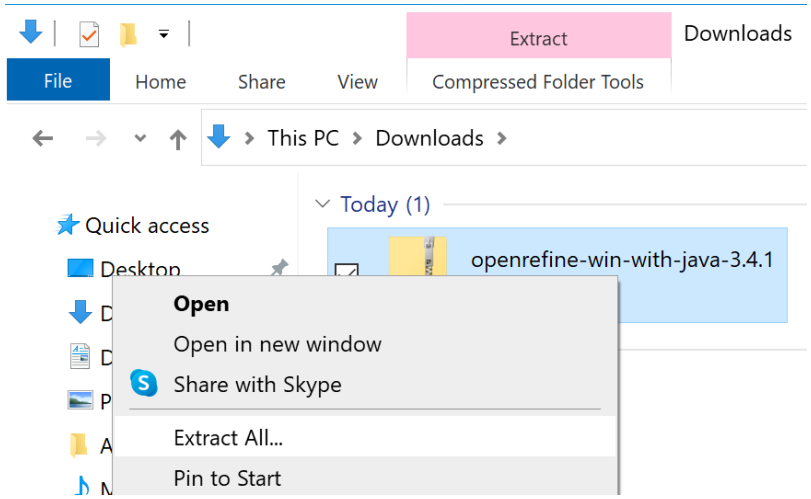
在微軟**Windows**上安裝

1. 下載 Windows kit **with embedded Java** 。選擇保存文件，而非打開它。
2. 找到下載好的檔案，點選右鍵選擇「解壓縮全部...」。解壓縮，並雙擊openrefine.exe
   (若打不開，則點選 refine.bat)。
3. 接著會跳出一個命令視窗(別關掉它)，稍待後，會出現一個新的web瀏覽器視窗，並顯示該應用程式。

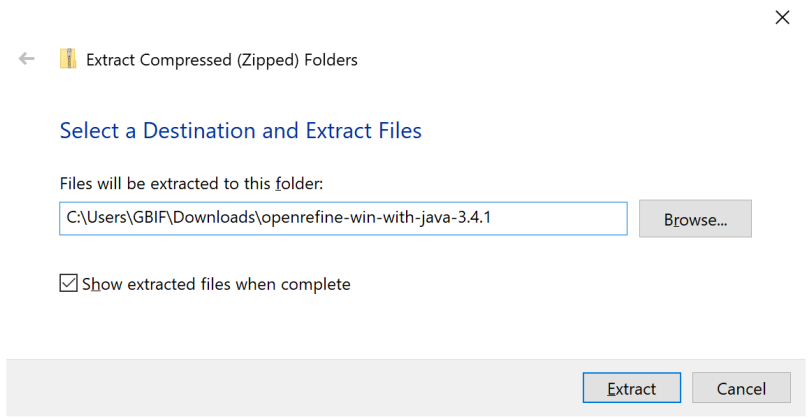
▼ 詳細說明__*Windows*系統(點選展開)

1. 

下載 Windows kit **with embedded Java** 。選擇保存文件，而非打開它。

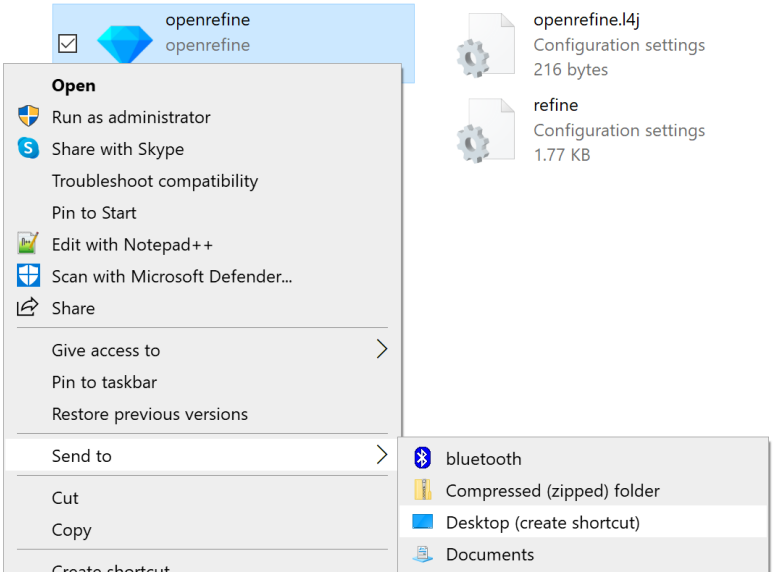
2. 

找到下載好的檔案，點選右鍵選擇「解壓縮全部...」

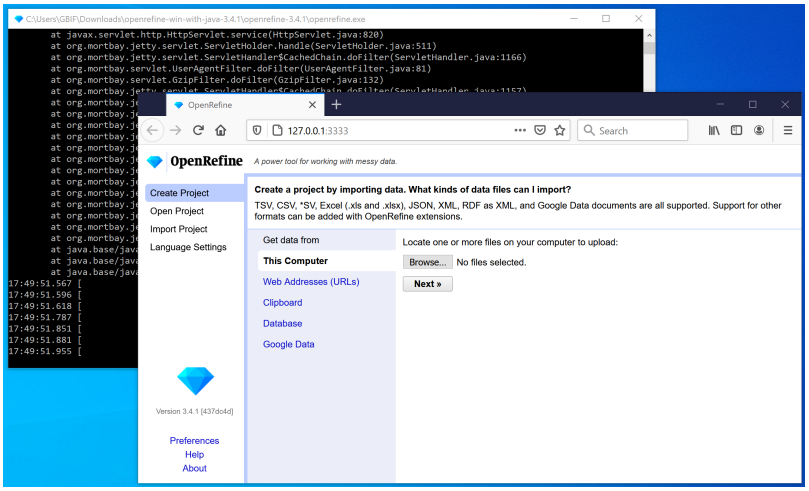
3. 

點選「解壓縮」

4.



找到解壓縮完的檔案，你可以點選「openrefine」後右鍵、選擇「傳送到 → 桌面(建立捷徑)」來在桌面上創建捷徑。接著雙擊點選開啟「openrefine」

5. 

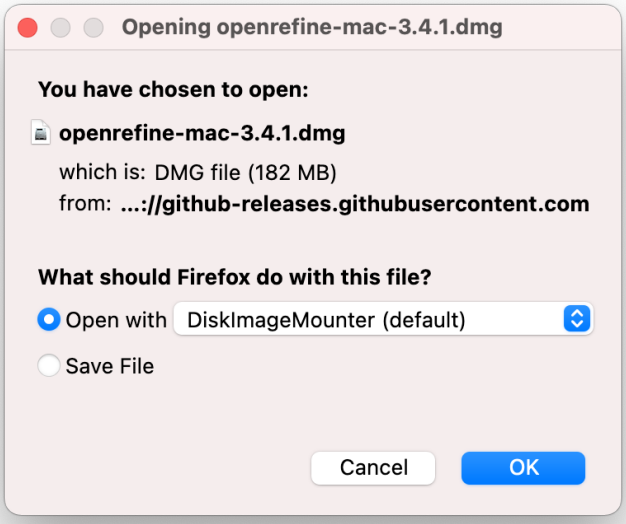
一個黑色的命令視窗開啟，短時間後瀏覽器會開啟，接著就能使用OpenRefine了。

在 **Mac** 上安裝

1. 下載 Mac kit 。
2. 下載且打開後、把圖示拖曳到應用程式資料夾。這裡你不需要額外分開安裝Java。
3. 雙擊點選它，就會跳出新的視窗來顯示應用程式。

▼ 詳細說明__*Mac*系統(點選展開)

1.



下載 Mac kit，並打開它。

2.

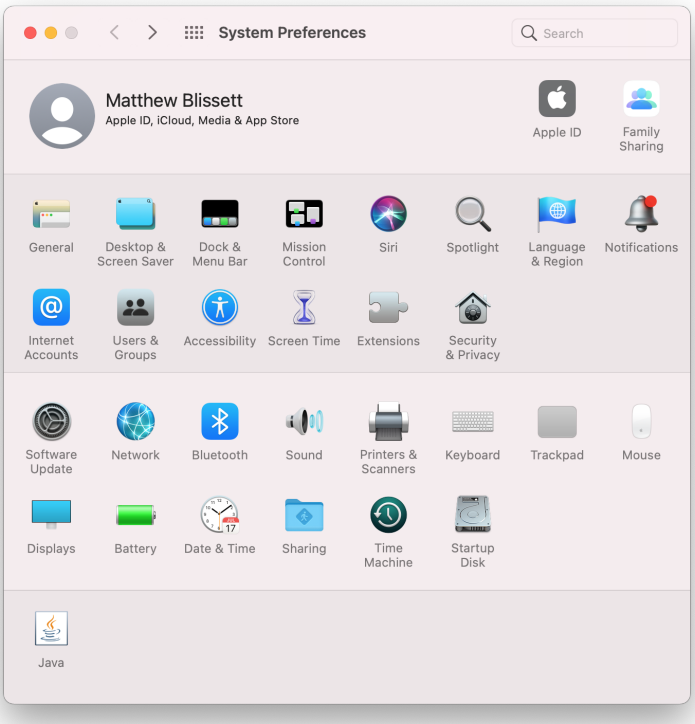
會跳出一個警訊，點選「OK」

3.

System Preferences

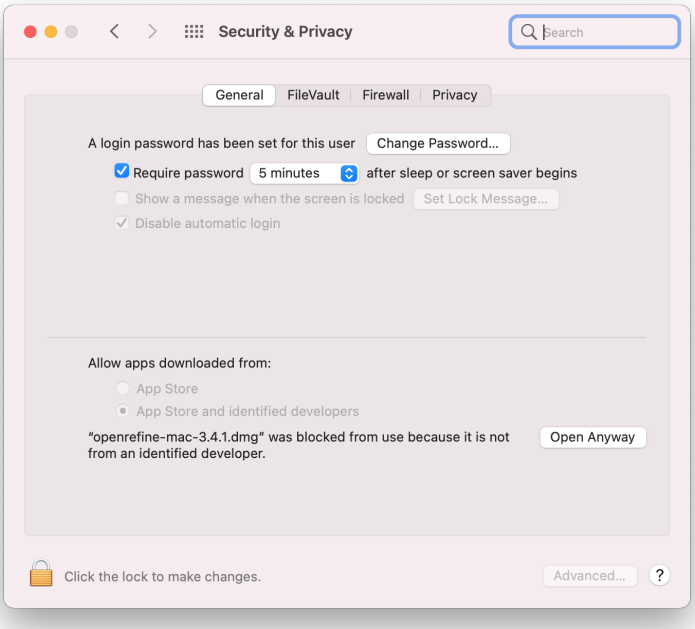
打開系統偏好設置

4.



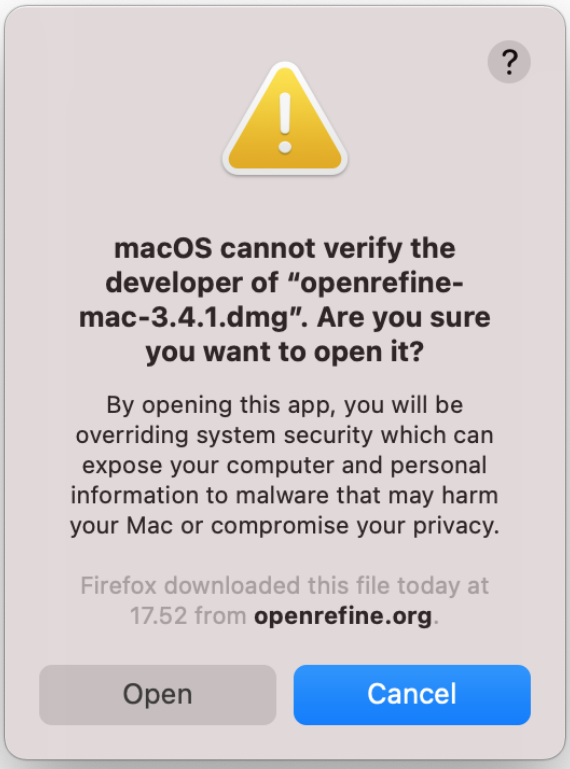
打開「安全與隱私性」

5.



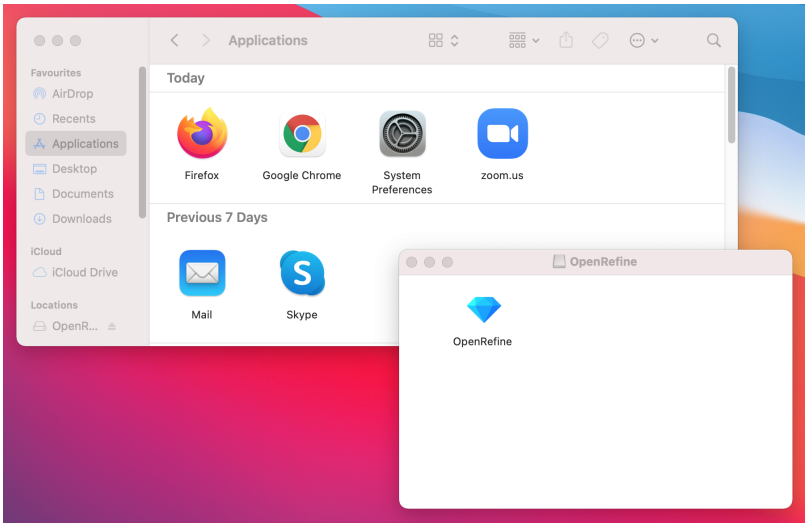
在按鈕上選擇「無論如何開啟」。

6.



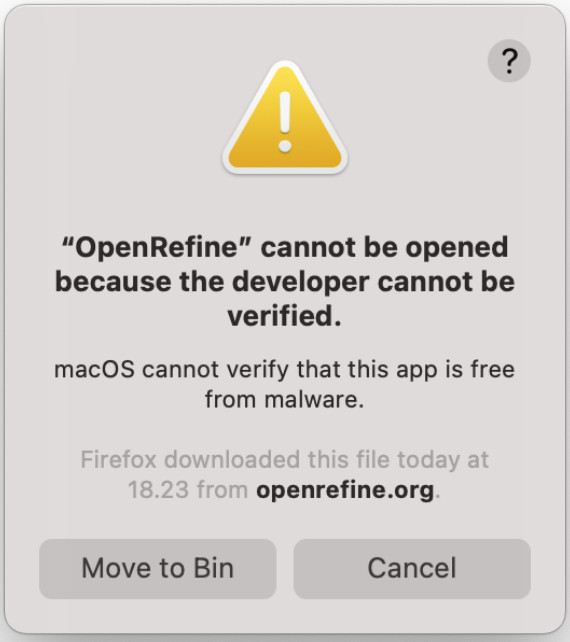
點選「打開」

7.



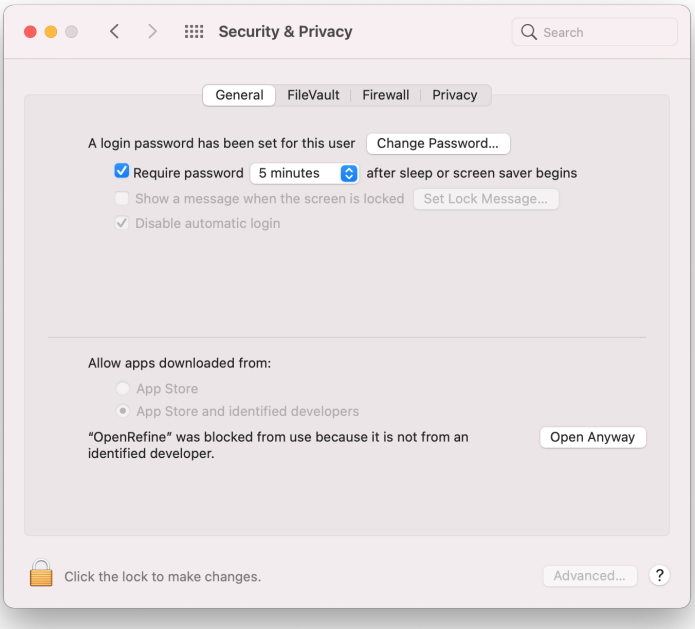
最終，應用程式存檔便打開了！把它拖曳到應用程式的資料夾中。

8.



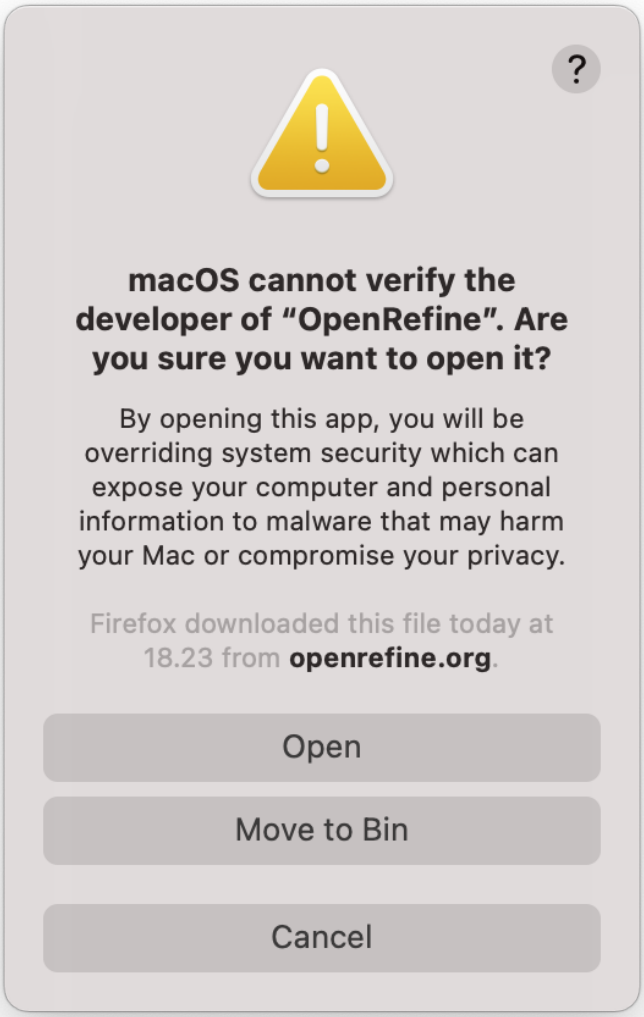
雙點OpenRefine 圖示，另一個安全性通知會出現。

9.



回到「安全性&隱私」，並再度點選「無論如何開啟」

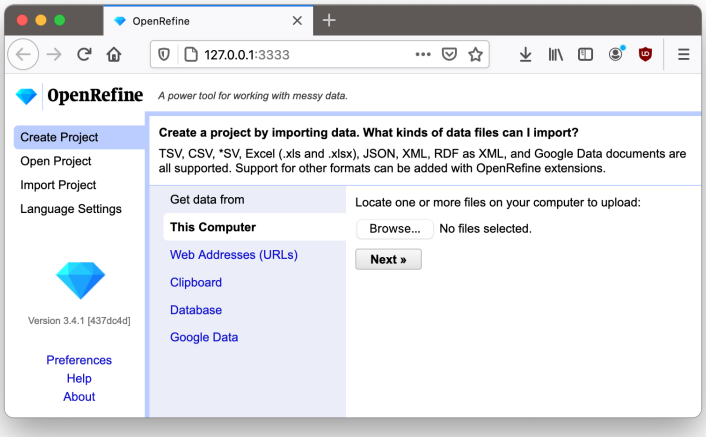
10.



(若要避免這些警示出現，OpenRefine 的開發者將需要付錢給Apple)

點選「開啟」。

11.



終於！應用程式開始運作了

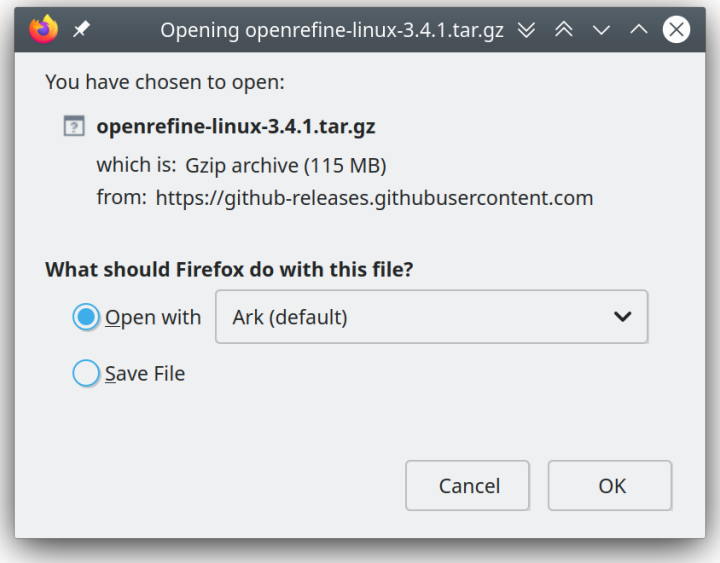
在 Linux 上安裝

1. 下載 Linux kit
2. 下載、解壓縮，接著輸入 ./refine 來開啟。這會需要Java有安裝在電腦上才能運行。

▼ 詳細說明__*Linux* 系統(點選展開)

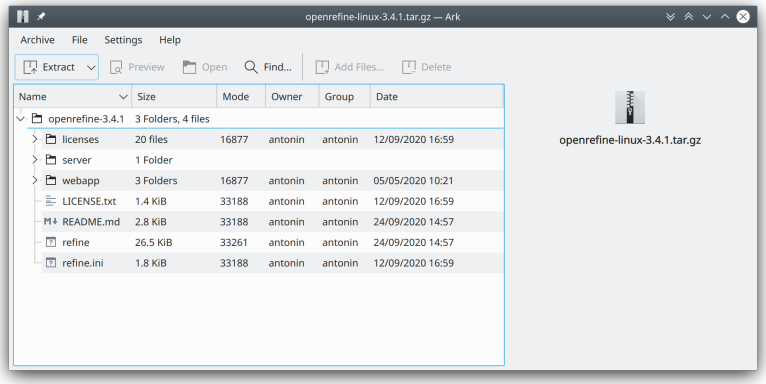
這些說明是寫給KDE(如： Kubuntu, SuSE)的，但過程與Gnome (如： Ubuntu, Red Hat, CentOS)很相似。

1.



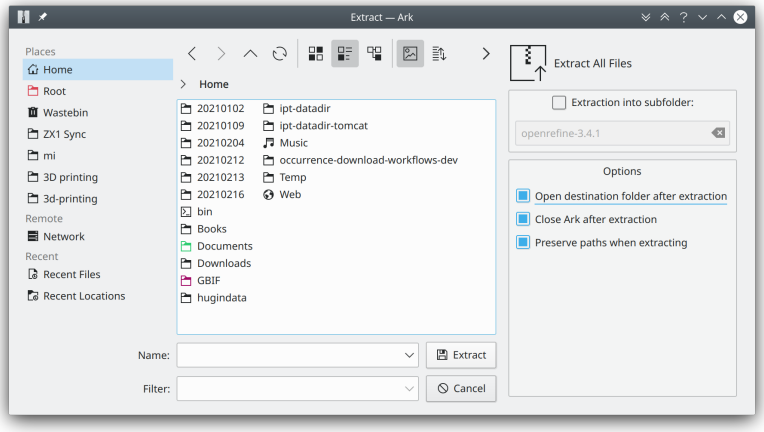
下載 Linux kit。 並打開它。

2.



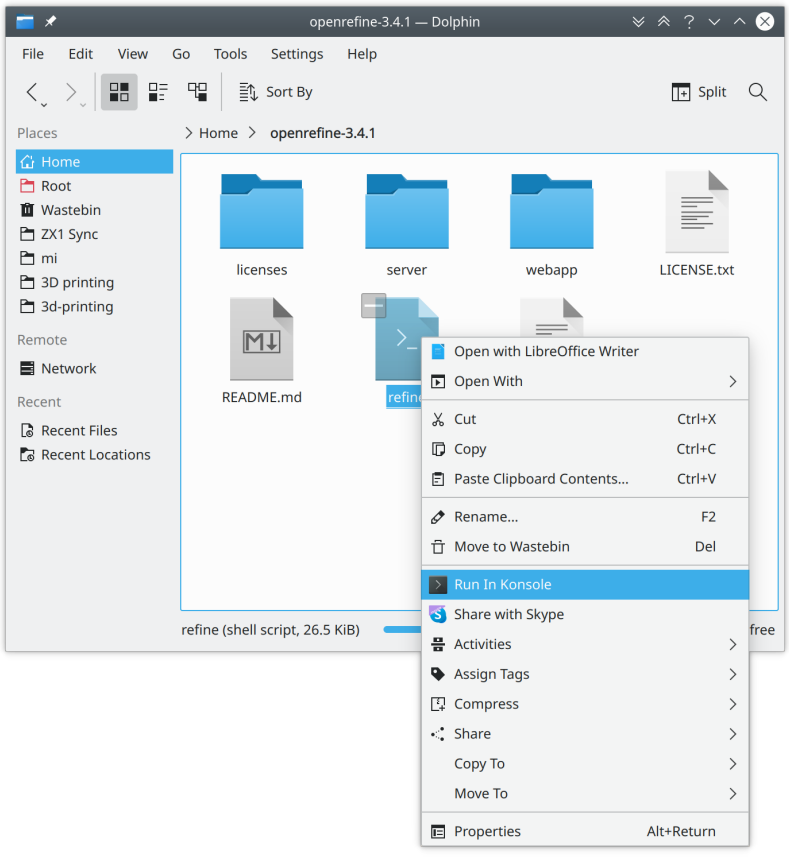
點選「解壓縮」來把下載的應用程式解壓縮。

3.



選擇適當的儲存空間。我這邊也一併選擇了"解壓縮完成後打開目的資料夾"及"在解壓縮後關閉Ark"。

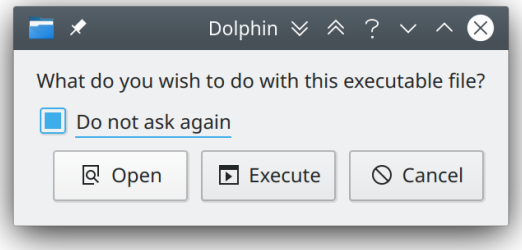
4.



右鍵點選「refine」且選「在命令視窗中運行(Run in Konsole)」。這是必要的步驟，來讓

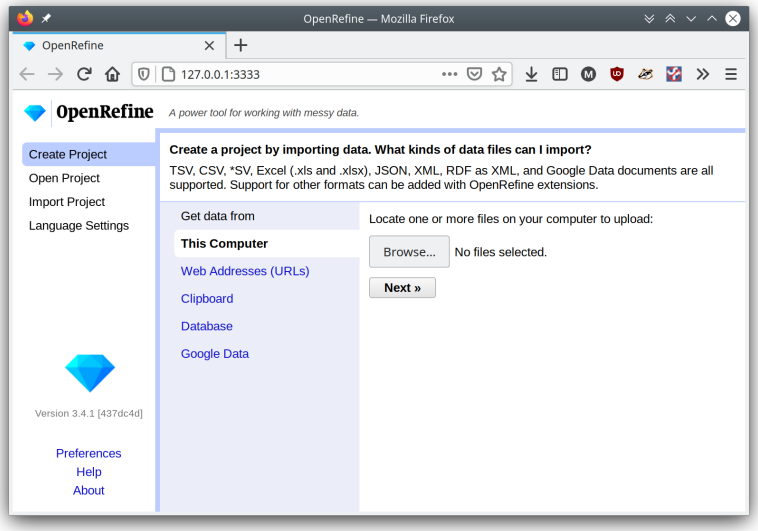
你在之後可以關閉Konsole視窗來安全地關掉OpenRefine 。

5.



確認你想執行下載的檔案

6.



OpenRefine開始運作中

基礎回顧

小試身手，回憶一下這節學到的概念吧！

1. 請在接下來的段落中填入對應的詞彙(資料庫、資料庫語言、資料庫程式)
   - 將操作資料的功能，在統一的介面中組合呈現的是 ＿＿＿＿＿＿＿＿＿
   - 將電腦中的數據、資訊結構化，並組織起來的集合是 ＿＿＿＿＿＿＿
   - 人類與電腦溝通所使用的是 ＿＿＿＿＿＿＿＿＿＿＿
2. 若你打開資料檔後看到以下情形，會判斷它為何種情況？

```
tre, ou ne pas tre, cest l la question.
```

- ☐ 沒事
- ☐ 檔案損毀
- ☐ 開啟檔案時採用錯誤的編碼

- [ ] 寄件者用了奇怪的的字體

3. 請為下方軟體配對填入正確的功能(資料獲取、資料管理、資料清理、資料發布)

    - 整合發布工具(Integrated Publishing Toolkit, IPT)
      _____
    - Specify
      _____ 及 _____
    - iNaturalist
      _____
    - OpenRefine
      _____

4. 請為下列填入正確的資料類型(二進位制 binary, 布林變數 boolean, 單浮點小數 float, 整數 integer, 常整數long integer, 文字text, 非結構化文字unstructured text)

    - 1236975
      _____
    - 01101111
      _____
    - We walked 5 miles down the road west from the post office in the center of town. We then went 2 miles north on a dirt path to the river. Then we continued west along the river for another 5 miles.
      _____
    - 1024
      _____
    - 29.0
      _____
    - Yes/No
      _____
    - 6 rabbits were observed
      _____

5. 下面哪些詞彙適合來形容「欄位名稱(field/column name)」？

    - [ ] 須被指定Assigned
    - [ ] 描述性Descriptive
    - [ ] 用來鑑別Identifying
    - [ ] 可讀Readable
    - [ ] 獨一無二Unique
    - [ ] 成為使用者的中介User-interface

6. 下面哪些詞彙適合來形容「欄位標籤(field label」？

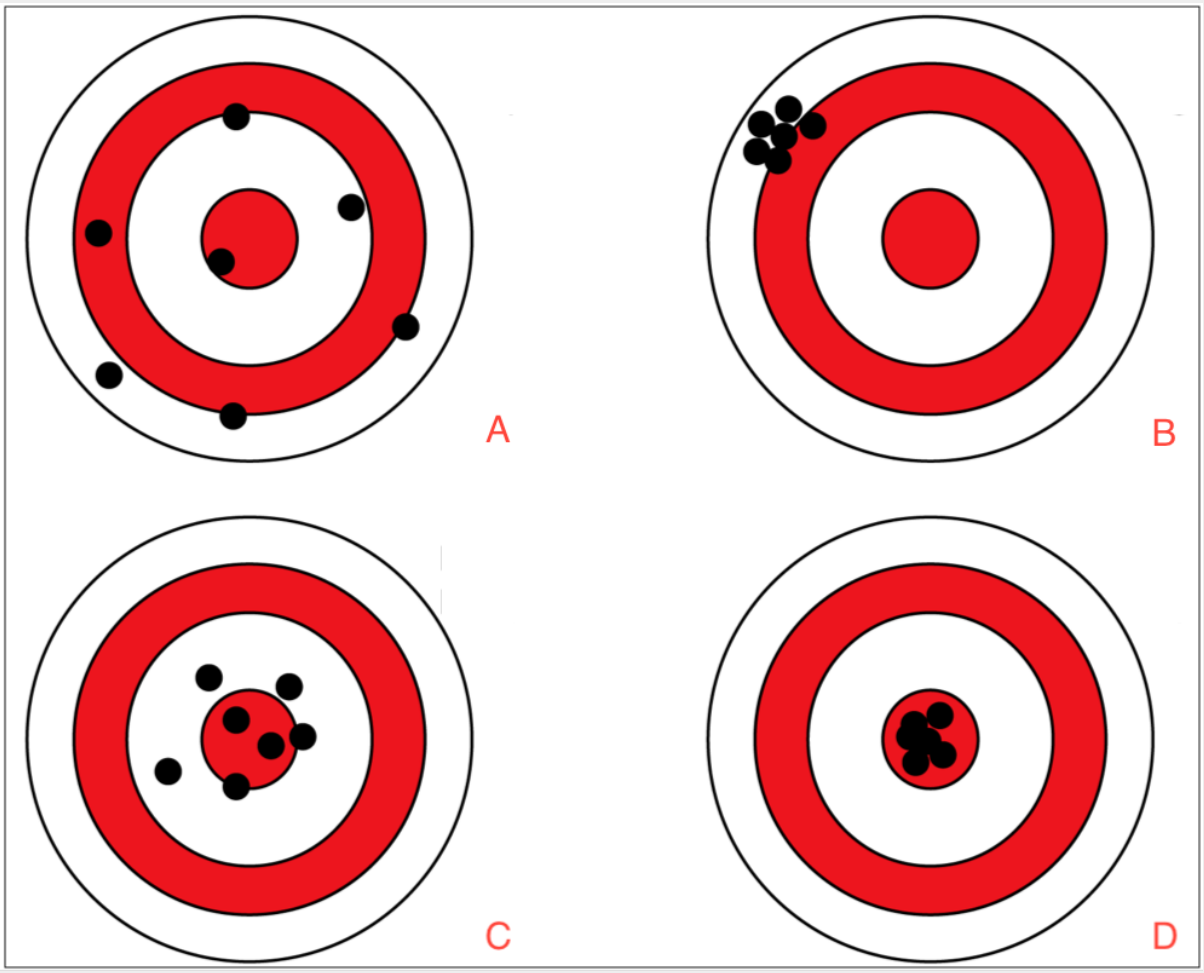
    - [ ] 須被指定Assigned
    - [ ] 描述性Descriptive
    - [ ] 用來鑑別Identifying
    - [ ] 可讀Readable

- [ ] 獨一無二Unique
- [ ] 成為使用者的中介User-interface

7. 請為下面敘述填入正確的資料結構 (列row, 行column, 表格table)
   - 集中了所有資料的是
     ——————————
   - 集中記錄同屬性資料的是
     ——————————
   - 將同筆數據的屬性資料綁在一起的是
     ————————

8. 誰決定了你資料的適用度(fitness for use)？
   - [ ] 博物館、部門主管
   - [ ] 用這筆資料研究、教育的使用者們
   - [ ] 該領域中的資料蒐集者
   - [ ] 將資料輸入資料庫的人

9. 請根據圖片填入適當的描述(A, B, C, D).



   - 高正確性,、低精準度
     ——————
   - 低正確性、高精準度
     ——————

- 高正確性、高精準度

  _____
- 低正確性、低精準度

  _____

10. 請在下列各個將資料集B併入資料集A的案例中，鑑別出是何種資料對應關係(0:1, 1:0, 1:1, 1:∞, ∞:1, ∞:∞，並非所有關係都會被使用到)

    - A、B中都存在的蒐集者欄位+ _____
    - 僅在B中存在的國家代碼

      _____
    - A中存在完整姓名欄位，而在B中拆成姓、名兩個欄位

      _____
    - A、B中都存在的ID欄位

      _____
    - 僅在A中存在的海拔紀錄

      _____
    - A中存在之日期欄位，而在B中被拆成年月日等不同欄位

      _____

11. 後設資料的重要是因為……(請選擇正確的敘述):

    - ☐ 它讓使用者能判斷是否符合用途
    - ☐ 它讓你能為每個物種出沒紀錄分享更精準的定位
    - ☐ 它讓人知道資料再利用的合法範圍、條約
    - ☐ 它能被應用在所有相關的材料，如：圖像、影片、或其他形式媒介上
    - ☐ 它能讓人知道該機構的展覽、與開放日

練習案例一、植物標本館的標本

請將自己帶入下列應用場景：

此案例是關乎於　規劃、資料獲取、資料管理與發布資料集　等部份的綜合練習。在此推薦先下載　exercise sheet　(MS　Word　345　KB)　，這樣方便在練習時邊做筆記。此案例的參考解法將在解答附錄中提供，此練習不會作為評分對象。

情境

地區植物標本館中的資料數位化專案

這是一個在瓜地馬拉蒐集的 裂瓣穀精草標本 Eriocaulon bilobatum Morong *by Rapid Reference Collection (RRC) | Field Museum of Natural History - Keller Action Science Center (licensed under* CC-BY-NC 4.0*)*

以下內容是作為資料數位化課程的基礎練習而設計的，其中構想、內容來自於 Alberto González-Talaván, Néstor Beltrán, Nicolas Noé 與 Sharon Grant 。使用到的資料則來自於實際存在的資料集(但有先因應本次練習而進行調整)，而相關的敘述皆是假想情境，且僅適用於教學目的。

描述

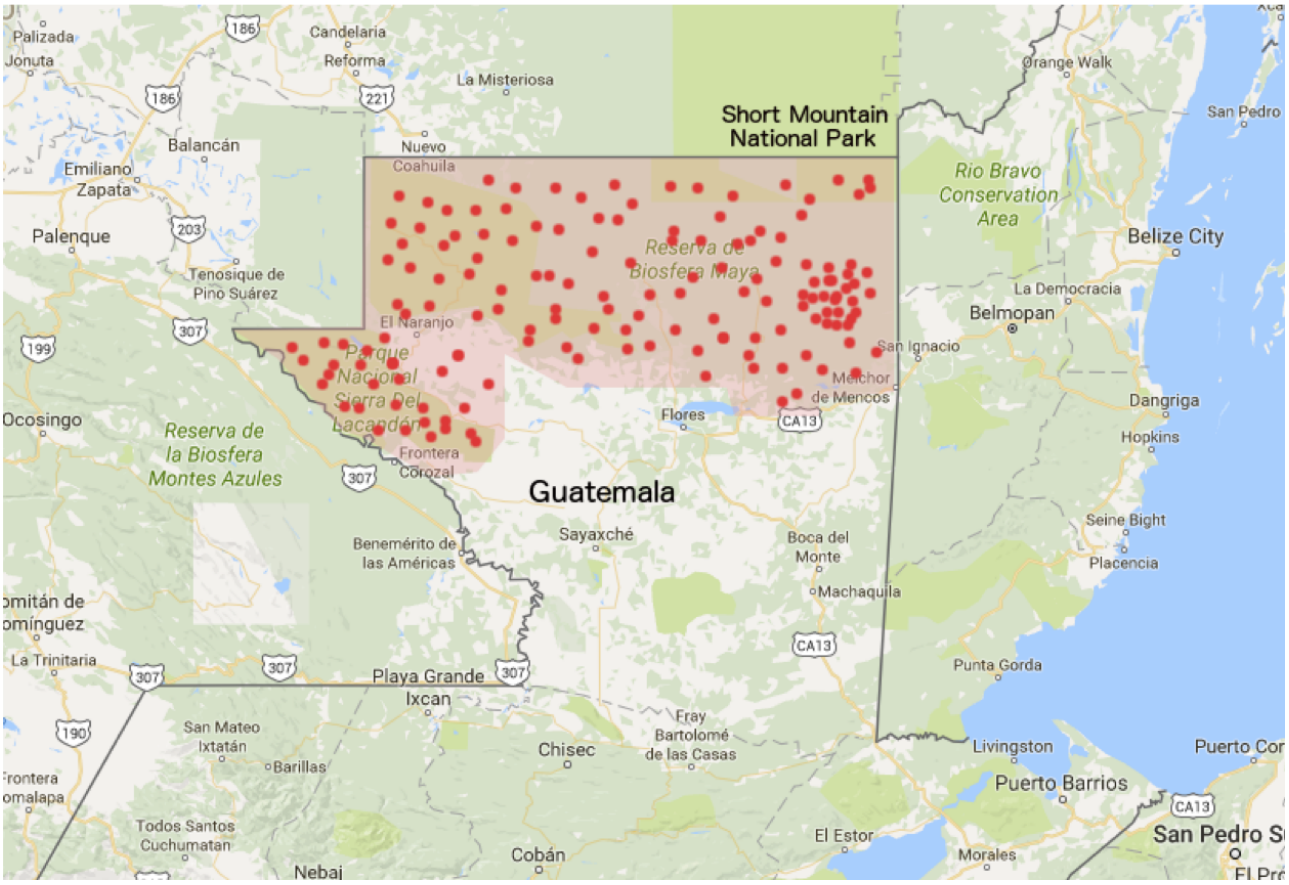
White Plains大學是一所積極新創、學科齊全、學術實力雄厚、辦學特色鮮明，在國際上具有重要影響力而廣為人知的瓜地馬拉高等教育機構，尤其在生物多樣性的研究更堪稱國家標竿。該校的植物科學系設有一中等規模的植物標本館，蒐羅該地區從20世紀中葉至今80,000多份的植物標本，其中不乏模式標本、或特有種標本等重要館藏。

當今，這些館藏主要由一位植物系統分類學教授管理，他負責策展工作、日常研究與教學。而部門管理員則負責植物標本室的日常庶務，如：購買紙張、標籤等耗材。系上的教職員工與學生偶爾會研究、更新標本的鑑定，且有兩位退休植物學家的定期志願來協助教授準備貸款。

該大學已設有一套網路檢索平台，可適用於圖書館館藏上。此功能由校內中央網路團隊在外部託管之伺服器上維護，而該平台目前尚無提供任何自然史標本之檢索。

植物生物學系主任最近獲得了為期兩年的5萬元美金資助，將用把館內植物標本影像、資訊的數位化開放。該團隊想藉機建立永久的資料數位化發布流程，以此提高植物標本館的能見度並繼續吸引贊助。

資料收集



植物標本館持有的標本約80,000份，採集於1960年至2015年，並集中於一個特定的生物多樣性熱點：矮山國家公園。由於樣本的交換、捐贈與數個進行中的研究計畫，館藏數量仍正在成長中。每個標本上的採集者、時間、日期、地點與分類資訊都有被記錄下來。至今為止，沒有進行過任何資料品質控管的手段，且從未無系統性地將館藏影像數位化過。

資料集說明

教授的電腦中有個非關聯式資料庫，能做為大部分標本的索引(但沒有涵蓋到所有標本)。任何標本的影像資料都只存在於當初拍下影像的研究者手上。

練習

此案例的各環節練習分布於對應的階段模塊中。

規劃

Exercise 1a-c

資料擷取

Exercise 2

資料管理

Exercise 3a-c

資料發布

Exercise 4

練習試算表

UC1-Herbarium-Exercise_EN.docx (MS Word 345 KB)

規劃

在規劃階段中，將會回顧關鍵的專案規畫階段，並學習建立可行的工作流程。接著，將會根據 USE CASE I 來創造理想的專案規劃、工作流程。你將確定其中的目標、任務、重要利益關係人、角色，並指派具體任務到各階段中。

資源盤點

在影片中，你將學習如何定義出會對專案產生影響的元素，並思考它們間的交互影響。若您無法觀看嵌入在課程頁面上的影片，可以點擊 download 下載至電腦中. (MP4 - 25.1 MB)

▶ https://www.youtube.com/watch?v=VRvUdMjd93c *(YouTube video)*

規劃

影片中(15:52)
將專注於任務、組織該項任務的方式，以便準備實際可行的方案及清楚明瞭的文件。若您無法觀看嵌入在課程頁面上的影片，你可以點擊 download 下載它 (MP4 - 26.2 MB)

▶ https://www.youtube.com/watch?v=uhhK6B2Vwls *(YouTube video)*

Exercise 1a-c

在這活動中，你將使用虛擬桌面平台。請看接下來的練習說明。

參考資料

- Role definitions
- Stage definitions

為了讓您能完成所有案例的規劃練習，我們提供了五個虛擬卡牌桌面連結。

由於它們為大家共享使用的空間，所以請您遵守下述守則，以方便彼此使用：

1. 用完後請自行清理。當您使用完卡牌桌面後，請清理至原本的樣貌。具體來說，就是將所有的卡片收回牌組，並移除所有新加入的卡片。這可以透過點擊每副牌下方的收起(recall)按鈕完成。
2. 請勿干擾他人。若您發現有人正在使用，請點選另外一個虛擬桌面連結。若您找不到空桌，請告訴我們 training@gbif.org ，這樣我們能進一步檢查哪些是無人看管的桌面(或為需要您開一張新桌面)
3. 請勿讓桌子無人看管。使用牌桌時，切勿長時間離席。可接受短時間(1、2小時)的離席，但請勿離席超過4小時。若您需要長時間離席，請利用eLearning platform 平台通知團隊、訓練員，並說明：所使用的連結、需要長時離席的理由、與預計返回的時間。
4. 善用螢幕截圖。我們無法保證您的卡片不會被來自工作坊、或您團隊的成員不小心重設位置——當卡片被召回牌組、重置後，是無法再度復原成先前的樣子的。若您想提交練習的螢幕截圖做為作業，請這樣做。

5. 訓練者保留重置牌桌的權利。若我們發現一個、或多個牌桌上有著超過4小時尚未收拾的卡片(有以表單、eLearning platform平台通知我們者除外)，我們將會清理以供他人使用。

卡牌遊戲間連結 (每個桌面將在新的分頁中開啟):

BLUE: https://playingcards.io/ndqppx

GREEN: https://playingcards.io/zpwe9j

ORANGE: https://playingcards.io/868jgh

PURPLE: https://playingcards.io/g6khb7

RED: https://playingcards.io/3pad4w

YELLOW: https://playingcards.io/w488nb

在此特別感謝 Jwalant Patel 和 Eric Ma 為本次練習找到並幫助建立此線上卡牌桌，以及Kate Webbink的美術專業能力。線上卡牌桌遊平台由by https://playingcards.io/ 提供。

Exercise 1a

請看 [練習案例一：情境] (若您還沒閱讀過的話)

透過卡牌選出最符合此專案大綱的的目標(goal)，接著選出需要進行的任務目標(task)。接下來，確定專案中可用的參與者、資源，並分配適合的角色卡給他們。最後，將這些分派給他們的利害關係人群體、隸屬關係上。

1. 一同查看目標卡(GOALS)，選擇並將適合於案例情境的目標卡放入其中。
2. 查看任務卡(TASK)。
3. 將任務卡指定到每個目標卡底下。
4. 辨別出在案例中提到的機構、人物，並記下他們。
5. 翻開隸屬關係卡(AFFILIATIONS)，並擺在桌上。
6. 查看利害關係人卡(STAKEHOLDER)，辨別在本次案例中提到者，而後決定他們的隸屬關係(AFFILIATION)
7. 查看角色卡(ROLE)，找出案例中提到者，並決定各個利害關係人(STAKEHOLDER)屬於何種角色。
8. 根據需要在卡片上做記錄。
9. 一旦分配了卡片，就拍照/截圖。
10. 用 exercise sheet 來提供你的答案

> 問題
>
> 是否有你認為對於完成專案至關重要，但卡片中遺漏的資源、目標呢？若有，請記在答案紙(answer sheet)上。

Exercise 1b

根據exercise 1a的利害關係人、目標分析，用工作階段卡(STAGE)來發展工作流。

1. 如果需要，可以重讀一次案例介紹。

2. 決定每個任務(TASK)該由哪個角色(ROLE)負責。
3. 查看工作階段卡(STAGE)，並適當地安排任務(TASK)。
4. 使用前面下載的練習表exercise sheet 來提供你的答案。

問題

- 工作流中有明顯的瓶頸嗎？如：有特定的角色/資源負擔了太多的任務嗎？
- 以各個利害關係人、角色的角度來看，哪些問題對成功的資料數位化來說是重要的呢？如：什麼是實際可交付的結果？以整體專案的時間表來說，他們是可行的嗎？
- 請整理好筆記，並照重要順序排好
- 如果有時間，你可以探索不同的組合，因為在不同的上下文中可能出現不同的場景，或甚至可以用自己專案的情況來做嘗試。

Exercise 1c

此練習應於課程以線上或現場小組活動形式進行時使用。

在這些練習後，各組的報告者將：

1. 補上此處缺少的利害關係人、任務，並說明應該被加入的理由。
2. 呈現兩個在小組內已找出最重要的議題/主題。

各組間可討論的方向：

- 彼此的工作流上有怎麼樣的相似、不同處？
- 是否有任何各組都共同面對的問題呢？

複習回顧

小試身手，回憶一下這節學到的概念吧！

1. PMBoK (Project Management Body of Knowledge, 專案知識管理流程群組)的五個步驟為何呢？
   - ☐ 規劃(Planning), 起始(Initiating), 監視與管制(Monitoring and Controlling), 執行(Executing), 結束(Closing)
   - ☐ 起始(Initiating), 規劃(Planning), 執行(Executing), 監視與管制(Monitoring and Controlling), 結束(Closing)
   - ☐ 起始(Initiating)，規劃(Planning)，執行(Executing)，結束(Closing)，監視與管制(Monitoring and Controlling)
   - ☐ 起始(Initiating), 規劃(Planning), 監視與管制(Monitoring and Controlling), 執行(Executing), 結束(Closing)
2. 所謂可交付的成果有哪些種類？(有數個正確答案)
   - ☐ 可被呈述的Stated
   - ☐ 隱含的Implied
   - ☐ 被估計的Estimated

- ☐ 直接的Direct
- ☐ 間接的Indirect
- ☐ 被推測的Guesses

3. 什麼是瓶頸？

- ☐ 推遲進程發展的阻礙
- ☐ 某人某物消失的空間
- ☐ 讓某人避免做某件事、或使其變得無法達成的問題或狀況

4. 哪些是資料數位化的的任務？ (有數個正確答案)

- ☐ 聯繫Affiliation
- ☐ 資料發表Publishing
- ☐ 影像化Imaging
- ☐ 地理參考Georeferencing
- ☐ 提高公眾關注Increased Public Awareness

資料擷取

在這個模塊中，你將學習與資料標準相關的觀念，尤其是達爾文核心(Darwin Core Standard)及其部分。你也將學習主要的生物多樣性資料類型，及如何在GBIF上分享它。最後，你將以資料擷取的視角來回顧資料品質的原則，並深入了解何謂資料品質與其一致性(特別著墨在：地理參考資訊、日期、名稱、及分類群之交叉檢查)。

資料標準與達爾文核心(Darwin Core)

這個影片中 (15:37), 你將了解你每天是如何與資料標準互動的。然後，您將了解 Biodiversity Information Standards, 包含 Darwin Core Standard ，此將在接下來的課程中持續使用。 若您無法查看鑲嵌於此的影片，可以點此下載觀賞 download (MP4 - 27 MB)

▶ https://www.youtube.com/watch?v=S02PJHPsRAs *(YouTube video)*

資料來源與類型

在這影片 (10:45)中，將查看能在GBIF上的資料類型 primary biodiversity data 。若您無法觀看嵌入在課程頁面上的影片，可以點此下載 download 並在電腦上觀賞。 (MP4 - 19 MB)

▶ https://www.youtube.com/watch?v=wKeOveydjsw *(YouTube video)*

問題

- 你的資料類型有不同於你原先所想像的嗎？
- 你工作上是使用怎樣的資料類型？

- 你將如何發表資料至GBIF呢(用哪個Core核心、Extension延伸表單)？

資料獲取、處理與品質

在此影片中(09:11)，你將探討何謂在資料獲取時的品質、與原則，特別是指自館藏標籤、田野筆記、或試算表來源獲取資料的場合。若您無法觀看遷入於此的影片，可點選 download 下載並觀賞。(MP4 - 19 MB)

▶ https://www.youtube.com/watch?v=QkDJIkmwBMA *(YouTube video)*

Exercise 2

在此練習中，你將演練資料的獲取來完成練習。你將開始使用達爾文核心標準 Darwin Core terms ，並決定哪些是你的組織/專案需要的資料欄位，並考慮在這之中的哪些將在後續的資料發布中分享出來。

請看 [練習案例一：情境] (若您還沒閱讀過的話)

想像你是那個被指定要把來自植物標本館表格中的資料轉錄出來的人

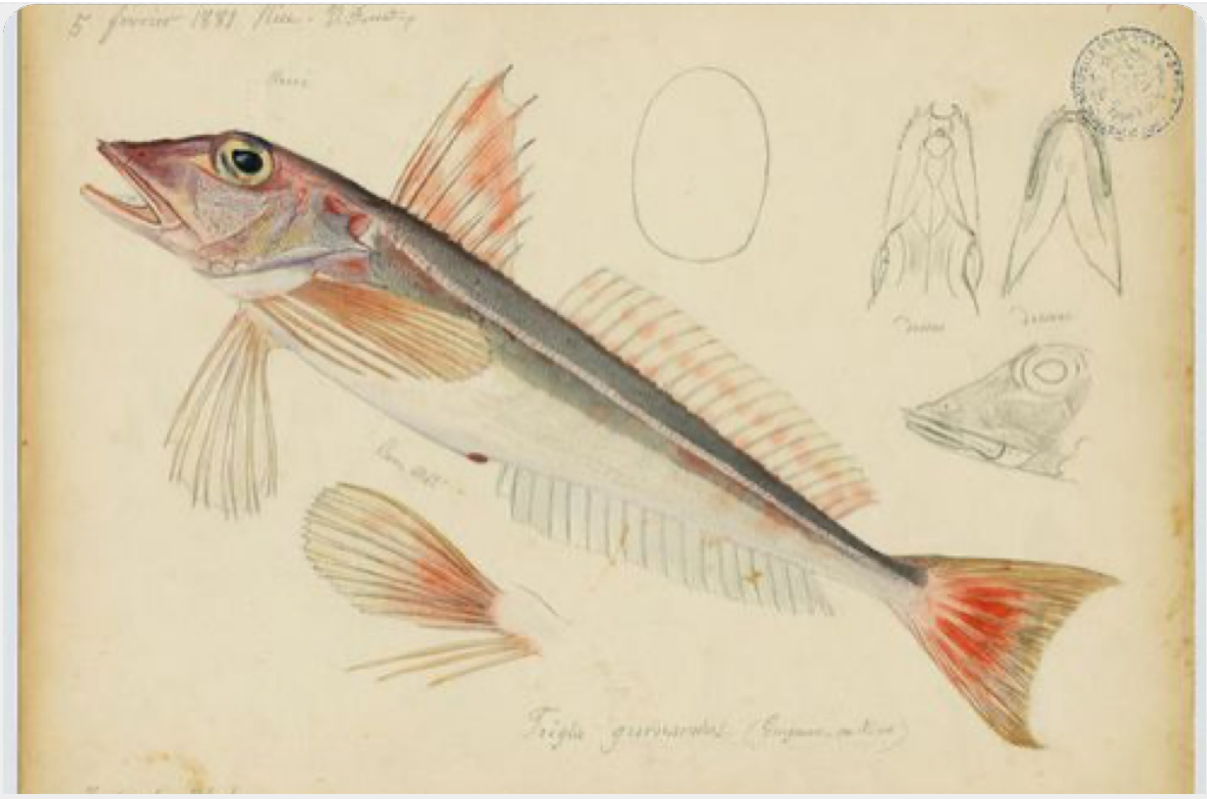
1. 下載 UC1-2-base-material.zip. (34.4 MB) 裡面共有10張圖片，其中每個標本各兩張，總計五個標本。植標館表格以西班牙文書寫(畢竟資料可能來自不同的管道、語言)，但你應該仍能認得在各項欄位中的資料。請記得使用每個標本各自的兩張影像來撰寫完整的資訊。
2. 下載試算表模板: UC1-2-occurrence-template.xlsx (57.3 KB)，用它來把照片上找到的資訊轉錄下來。
3. 使用前面下載的練習表exercise sheet 來提供你的答案。

你或許會加入更多欄位，因為你可能能夠找到比模板預期規劃中更多的欄位資訊。

複習回顧

小試身手，回憶一下這節學到的概念吧！

1. 你會選擇哪種資料型態給一個魚類分類學的蒐藏呢？



Eutrigla gurnardus (Linnaeus, 1758) | *Muséum d'histoire naturelle de Nice*

- ☐ 物種出現紀錄occurrence
- ☐ 物種名錄checklist
- ☐ 調查事件sampling event

2. 你會選擇什麼資料型態給一個關於入侵種的清單呢？

Water hyacinth (Eichhornia crassipes) observed in Bourail, New Caledonia, where it is an introduced and invasive species by GRIIS. Photo by gérard (2016) licensed under CC BY-SA 2.0

- ☐ 物種出現紀錄occurrence
- ☐ 物種名錄checklist
- ☐ 調查事件sampling event

3. 你會選擇什麼資料型態給環境衝擊調查下所得的植物相、動物相資料呢？

   這份環境衝擊評估調查是由專家完成，目的是為了評估一特定區域在人類活動影響(如：道路施工、風機發電、採礦作業、建築…等等)之前、之下、之後的的生物多樣性、棲地類型(biotope)。

Entomologist chasing butterflies *by Matthieu Gauvain (CC-BY-SA)*

- ☐ 物種出現紀錄occurrence
- [] 物種名錄checklist
- ☐ 調查事件sampling event

4. 你會為鳥類的追蹤資料選擇怎樣的資料集型態呢？

   這份鳥類追蹤紀錄是利用特殊裝置(如：安裝於鳥身上的GPS追蹤器)紀錄而成，也是此能讓科學家得幾追蹤牠們的遷徙路徑、築巢地點。

Griffin vulture *observed at Gamla Nature Reserve by גיזונית - MinoZig (CC0)*

- ☐ 物種出現紀錄occurrence
  - [] 物種名錄checklist
  - [] 調查事件sampling event

5. 你會為捕蟲陷阱資料選擇怎樣的資料集型態呢？

Insect trap *by miheco (CC-BY-SA)*

- ☐ 物種出現紀錄occurrence

[] 物種名錄checklist

☐ 調查事件sampling event

6. 你會為國家公園管理資料選擇怎樣的資料集型態呢？

特定保護區域(如：國家公園、或小的自然保留區)管理所需要的資料十分多樣、且具有不同的資料來源：植生調查、動物個體追蹤、巡守員或警衛的觀察、甚至是「公民科學」資料或從社群媒體圖片中可推論出來的資料。

Sri Lankan elephants *observed by pen_ash.*

☐ 物種出現紀錄occurrence

☐ 物種名錄checklist

☐ 調查事件sampling event

7. 你認為公民科學中生物閃電戰(bioblitz)應該要用什麼資料集型態呢？

公民科學資料有時會透過主題式的田野工作日來蒐集獲得，也就是生物閃電戰(bioblitz)。志願者們通常聚集於特定區域，並花上一天試圖觀察、鑑別儘可能越多物種。

每個參與者產出的資料，都會被彙整進整個公民科學專案之中、或資料管理工具內。

Looking for birds with park staff *by US National Park Service (authorized reuse on google image search)*

- ☐ 物種出現紀錄occurrence
- [] 物種名錄checklist
- ☐ 調查事件sampling event

8. 你會為一地區物種名錄選擇哪種資料集型態呢？

Black rhino *observed at the Magdeburg Zoo in Germany by Mani300*

- ☐ occurrence
- ☐ checklist
- ☐ sampling event

資料管理

在這個模組中，你將回顧資料管理(尤其是資料清理與標準化)的主要概念、最佳操作以及相關工具。

資料管理的原則

在這影片 (09:49)中，將回顧一些透過資料清理過程中善資料的重要原則。若您無法觀看嵌入在課程頁面上的影片，可以點此下載 download 並在電腦上觀賞。 (MP4 - 16.6 MB)

▶ https://www.youtube.com/watch?v=4ijm1cJeVHE *(YouTube video)*

資料管理工具

在這影片中 (06:42)， 你將學到一些能用來增進資料品質的工具。若您無法觀看嵌入在課程頁面上的影片，可以點此下載 download 並在電腦上觀賞。 (MP4 - 10.3 MB)

▶ https://www.youtube.com/watch?v=Ru3vWiYU3gw *(YouTube video)*

Exercise 3a-c

這個練習中，你將進行技術性驗證與一致性檢查，以不同工具來增進資料集外，且學習如何使用 OpenRefine 。

請看 [練習案例一：情境] (若您還沒閱讀過的話)

你的機構隸屬於「國際禾本植物協會(Global Poales Association, 下稱GPA)」的其中一員。此協會已經募得了發表最新版植物誌所需的資助，並要求你的標本館去參與、並提供該目植物任何可能有的高品質紀錄資訊。該目在你的館藏中十分豐富，因此你認為你能很好地對此有所貢獻。

Exercise 3a

資料驗證檢**查**

此練習中，我們將專注在技術性錯誤，並練習執行基本的驗證檢查來找出技術性錯誤。請參考 [Validation checks] 來得到更多關於錯誤的種類資訊。

1. Download UC1-3ab-data-cleaning.csv. (207.5 KB)
2. 將CSV以Excel wizard匯入Excel中。可見 Excel-tips-EN.pdf (PDF, 7 MB) 來獲得符合您操作系統(Windows, Mac, Linux)的更多說明。
3. 找到、手動更正錯誤。
4. 使用前面下載的練習表exercise sheet 來提供你的答案。

Exercise 3b

其他資料管理工具

GPA協會給了一份資料品質的要求清單要你驗證：

- 所有植物名稱(全名)都拼寫正確
- 所有植物名都屬於該目
- 所有紀錄都有座標
- 所有座標都在上面表示的國家內、並轉換至十進位制
- 所有日期資料都在適當的欄位內，並且格式為YYYY-MM-DD

錯誤的三種種類分別是：

- 命名錯誤
- 格式錯誤
- 地理資訊錯誤/離群值

1. 請參考 [Helpful tools] 來完成整個練習。可以使用的工具不受限於這些，您可以使用任何想要的工具。
2. 使用上個練習中的相同檔案
3. 只對 Eriocaulaceae (穀精草科)的紀錄進行更正(對此你可能會要篩選資料)
4. 更正那些在前個練習中(exercise 3a)找到的資料集錯誤，使用你選擇的工具，並將你做出的變更在練習表中記錄成檔。
5. 若您有空的話，把整個檔案都做一遍。
6. 使用前面下載的練習表exercise sheet 來提供你的答案。

Exercise 3c

在這影片中 (03:27)，你將學習 OpenRefine。你可以用 OpenRefine來標準化、並增進你的資料集品質。若您無法觀看嵌入在課程頁面上的影片，可點此下載 download 至電腦上觀賞 (MP4 - 3.8 MB)

▶ https://www.youtube.com/watch?v=_YFw_bfwc3Y *(YouTube video)*

OpenRefine

在此練習中，我們使用OpenRefine 來改善資料集的品質。將使用的是內建的基本功能、既存的網路服務、及正規表示式。

1. Download UC1-3c-open-refine.csv. (207.5 KB)
2. 下載並完成連結中的練習 in OpenRefine-Exercise3c-EN.pdf. (PDF, 1.1 MB)。亦可於此取得不同語言版本 French 及 Spanish.
3. 使用前面下載的練習表exercise sheet 來提供你的答案。

練習的小技巧

資料驗證檢查

技術性錯誤**Technical errors** 相對簡單、容易被自動化地檢查出來，檢查資料的完整性。這也指那些錯誤的輸出格式、資料映射、欄位滑動(如：整行向右滑動一行)、及來源的資料缺失。

- 完整性**Completeness**: 所有的資料與後設資料是否齊全——是否所有欄位都在、且無缺失？
- 該有的限制**Bounds**: 比如日期應當於1至31之間(依月份而定)
- 資料類型**Data type**: 舉例來說，填在日期欄位中的是否是單一日期、或數字呢？
- 資料格式**Data format**: 舉例來說，日期格式會是如 01/01/2010 還是 01/Jan/10？

一致性錯誤**Consistency errors**

把真實世界的規則套用在資料上。這可能代表有些錯誤是來自於舊紀錄的資料輸入、過去的錯誤轉錄、或後處理；而有些規則應用上十分複雜、且 需要參考資料集才能檢驗 ——如：一份關於已知的蒐集者及其蒐集習慣的清單。這些規則可以從資料使用者、分析師中得到。

- 分類上的**Taxonomic**: 如，若鑑定至種，則得有屬名+種小名等完整的學名。
- 流通性**Currency**: 檢查採集、鑑定、更新及數位化的日期是否一致？
- 異常值**Outliers**: 去找出異常值，但記住不是所有的異常值都一定是錯誤。比如：不同於已知的物種分布區域、或已知的環境範圍(但請記住，比起是座標的錯誤異常，這個異常值更可能是被錯誤鑑定至此而產生的結果)。
- 地理上的**Geographic**: 這些座標有落於對應的地點、地區內嗎？比如：是否有任何出現在海上的陸生紀錄、或出現在陸地上的海洋紀錄？
- 採集模式**Collecting patterns**: 此筆出現紀錄的細節是否符合這個組織、採集者他們已知的採集模式呢？ 是否有任何紀錄是在採集者死亡後新增的(或者，這可能是另一位有著同名的不同採集者)？舉例如：是否有任何哺乳類紀錄出現在賞鳥團體的紀錄當中？
- 正確性與精準度**Accuracy and precision**: 具體來說，是否有任何具地理參照的紀錄在精確的定位之前，就有著非常高的GPS精準度、正確性？
- 採集方法**Collecting methods**: 不同採集方法(如：穿越線法、區域調查法)具有特定的特徵。這些記錄呈現出來與提供出來的方法是一致的嗎？

實用工具

- **GBIF Name Parser**: https://www.gbif.org/tools/name-parser
- **Global Names Resolver**: http://resolver.globalnames.org
- **Catalogue of Life name match**: https://data.catalogueoflife.org/tools/name-match
- **TNRS**: https://tnrs.biendata.org/
- **WoRMS**: https://www.marinespecies.org/aphia.php?p=match
- **InfoXY**: http://splink.cria.org.br/infoxy?criaLANG=en
- **Georeferencing Calculator**: http://georeferencing.org/georefcalculator/gc.html
- **Canadensys coordinate conversion**: http://data.canadensys.net/tools/coordinates
- **Canadensys date parsing**: http://data.canadensys.net/tools/dates
- **Google Maps**: https://maps.google.com/

複習回顧

小試身手，回憶一下這節學到的概念吧！

1. 為什麼最好要清理你的資料呢？

   - ☐ 讓它們盡可能的方便使用
   - ☐ 達到你對資料品質的目標
   - ☐ 資料應該由使用者自己清，而非提供者

2. 如何建構自己資料清理的工作流呢？

   - ☐ 自己工作，只有自己最了解資料集
   - ☐ 詢問你的專家同事們
   - ☐ 以機構的等級切入，對焦建立資料品質的工作流

3. 下列何者是最好的：

   - ☐ 預防錯誤發生
   - ☐ 盡速更正你在資料集、資料庫中找到的錯誤
   - ☐ 不清理錯誤但將它們記錄下來，因此使用你資料的人們可以知道錯誤在哪裡

4. 資料品質是誰的責任？

   - ☐ 記錄此筆資料的人
   - ☐ 資料抄寫員
   - ☐ 資料庫管理者
   - ☐ 所有參與到資料管理的人
   - ☐ 用你資料的人
   - ☐ GBIF

5. 下列哪些工具可以用來清理你的資料？

   - ☐ Excel 和其他 spreadsheets的管理工具
   - ☐ OpenRefine
   - ☐ 你的資料庫軟體
   - ☐ 線上工具，如：Scientific Names Resolver 、 Google Maps

資料發布

這個課程模組中，你將學到資料集發布的概念，包含：IPT的使用、核心與衍生資料表單、授權的重要性、後設資料、必填欄位與資料集的擁有管理。

資料發布概念

在這影片中 (11:45)，你將學習資料集發布的相關概念，並且會學習如何使用IPT(Integrated Publishing Toolkit 資料整合發布工具，https://www.gbif.org/ipt[IPT^])的概要。若你無法觀看嵌入在頁面上的影片，你可以點此下載 download 至電腦上觀看 (MP4 - 20 MB)

▶ https://www.youtube.com/watch?v=b900d9ukjSQ *(YouTube video)*

IPT 介紹

在這影片中 (06:56)，你將看到IPT工具的資料發布介面的相關介紹。若你無法觀看嵌入在頁面上的影片，可以點此下載 download 至電腦上觀看 (MP4 - 8.7 MB)

▶ https://www.youtube.com/watch?v=gHXsaN_JWeI *(YouTube video)*

Training IPT installations IPT課程訓練測試站

若你沒有拿到登入用的帳號，請聯絡training@gbif.org，你將會拿到一組訓練用的課程帳號密碼。

https://training-ipt-a.gbif.org/

https://training-ipt-b.gbif.org/

https://training-ipt-c.gbif.org/

IPT 工具示範

在這影片中 (24:16)， 你將學到如何用IPT工具發布資料集。若您無法觀看嵌入在課程頁面上的影片，可以點此下載 download 並在電腦上觀賞。 (MP4 - 52.6 MB)

▶ https://www.youtube.com/watch?v=eDH9IoTrMVE *(YouTube video)*

Exercise 4

此練習中，你將使用IPT工具來發表一個物種出現記錄的資料集。

請看 [練習案例一：情境] (若您還沒閱讀過的話)

資料集發布**Data publishing**

在清理完禾本目的資料集後，團隊認為把資料發佈到GBIF上是提升可見度的好方法，因此你被要求主導資料集發布的工作。

1. 在這臉上，你需要一個 course IPTs 的課程用IPT帳號。若你還沒有取得帳號，請聯絡 training@gbif.org，而你將獲得一組課程用IPT工具的帳號密碼。
2. 下載 UC1-4-poales-publishing.csv. (233.5 KB)
3. 使用對應的IPT測試站，並發布此檔案。
4. 使用前面下載的練習表exercise sheet 來提供你的答案。

複習回顧

小試身手，回憶一下這節學到的概念吧！

1. 資料發佈是什麼意思呢？

- ☐ 把你清好的csv檔輸出，分享給你的同伴
- ☐ 寫個關於你資料集的文章，其中並描述資料蒐集、獲取、及清理的操作方法
- ☐ 讓你的生物多樣性資料集能以標準的格式被公眾獲取

2. 什麼是IPT？

- ☐ 幫助你管理、更正資料用的工具
- ☐ 用來發布資料到GBIF上的工具
- ☐ 協助你撰寫資料論文(Data paper)的工具

3. 在資料集發布上，GBIF推薦哪種的創用CC標章(Creative Commons licences)、授權協議呢？

- ☐ CC-BY、 CC-BY-SA、 CC-BY-ND
- ☐ CC0、 CC-BY 、CC-BY-NC
- ☐ CC0、 CC-BY 、CC-BY-SA

4. IPT選擇資料來源時，可選的三個核心表單(Core)為哪些？

- ☐ Metadata Core, Occurrence Core, Multimedia Core
- ☐ Taxon Core, Collection Core, MeasurementOrFact Core
- ☐ Occurrence Core, Taxon Core, Event Core

5. 一個資料集能附有幾個額外表單呢？

- ☐ 零個
- ☐ 一個
- ☐ 皆可，根據需求而定

學員考核與認證

在此課程模組中，你將了解用來考核學員、認證的標準。

在上完課程並通過作業考核後，便有機會在 Open Badge 中得到官方的相關認證。

An overall score of 2.5-2.9 earns a BASIC Biodiversity Data Mobilization badge

An overall score of 3.0-4.0 earns an ADVANCED Biodiversity Data Mobilization badge

參與者需要以英文繳交Use Case II(從兩個選項中擇一)及 Use Case III ，而他們都將根據課程的評量指標進行評分。這些評量指標定義了在每個學習目標上的技能表現。

重溫一次這些評量標準，來確保了解即將進行考核認證的技能。

規劃階段的評量標準

規劃階段

技能	基礎表現 **1**	進步中表現 **2**	熟練表現 **3**	傑出表現 **4**
A.了解生物多樣性資料流通計劃中的不同元素	能夠了解部分資料流通計劃中所需的角色和任務，但無法區分每個角色執行的具體任務，或將角色混淆。	了解資料流通計劃中所需的許多角色和任務，但仍缺乏對他們之間關聯與互動的認識。	了解資料流通計劃中所需的關鍵任務和角色以及他們如何相互作用。	能夠辨別針對特定情境可能需要的額外角色與任務。
B. 將資料流通計畫的不同元素應用於特定機構背景（如自身機構）的能力	能夠理解部分任務與角色與特定機構情境的關聯性，但在情境稍有變化時會感到困難，即使是稍微偏離使用的通用參考資料時。	能夠理解大多數任務與角色如何應用於特定機構情境，但在情境不同於通用參考資料時仍有不足。	即使情境與通用參考資料不完全一致，能夠辨別所有相關任務與角色並理解如何將他們應用於特定機構情境。	能夠有創意地應用資料流通計劃的不同元素，並結合通用參考資料中未提到的新元素。參考資料
C. 撰寫/編寫清晰計劃文件的能力	能夠收集計劃所需的初始素材（如要點），但難以將其轉化為完整敘述。	能夠構建基本工作計劃，但其元素之間缺乏連貫性。	能夠撰寫包含所有相關要素的連貫且清晰的工作計劃文件。	能夠撰寫簡潔的計劃，包括有效摘要，並且層層遞進地提供細節。
D. 評估生物多樣性資料流通計劃的能力	能從給定範例中辨別出一般資料流通計劃的少數要素，但難以將這些要素與具體任務和角色配對。	當計劃內容與參考資料相同時，能辨別出多數流通計劃要素，並將其與任務和角色配對。但在評估計劃某些要素的可行性時仍有困難。	能夠辨識資料流通計劃中的所有現有要素，並能找出計劃中的不足、重複工作和不一致性。能評估個別組成部分的品質。能評估整體計劃的潛在成功點和弱點。	能夠針對計劃中發現的問題提出解決方案。

資料獲取的評量標準

資料獲取

技能	基礎表現 **1**	進步中表現 **2**	熟練表現 **3**	傑出表現 **4**
A. 能夠辨別可從生物多樣性資料來源中摘錄的數位資料類型（即可透過 GBIF 網絡發布的資料	只能從常見的生物多樣性資料來源中辨別最明顯的資料類型（例如：自然史典藏標本的出現紀錄）。對使用 GBIF 進行線上發布的潛力理解有限。	通常能正確辨別至少一種可從常見資料來源摘錄的數位資料類型。但難以判斷哪些資料類型目前可以使用 GBIF 發布。	總是能辨別一種（或多種）可從常見資料來源摘錄的數位資料類型。能判斷這些類型中哪些目前可以使用 GBIF 發布。	總是能辨別一種或多種可從常見和不常見資料來源摘錄的數位資料類型。能判斷這些類型中哪些目前可以使用 GBIF 發布，哪些正在討論中。能判步用於發布這些資料類型的核心資料集以及延伸資料集。

技能	基礎表現 1	進步中表現 2	熟練表現 3	傑出表現 4
B. 將生物多樣性資料來源中的相關資訊摘錄到符合國際標準的簡單資料結構（例如：試算表）的能力	只能從資料來源中摘錄明顯的資訊（例如：將所有地理資訊作為單一單位）。對記錄生物多樣性資料的現行標準了解有限。	能從資料來源中擷取數個但非全部資訊項目，並能將其分解為有意義的片段。對最常見的標準（例如：DwC）和這些標準中最常用的資料欄位有基本認識。	能辨別資料來源中所有有價值的資訊，並將必要元素摘錄到標準資料結構中（例如：基於簡易 DwC 的試算表）。能辨別缺少的資訊，並從現有資訊推斷出來（例如：從省份資訊推導出國家名稱）。	能辨別複雜資料來源中所有有價值的資訊，並將其分解為有意義的片段，然後直接轉換為國際標準格式。能辨別來源中缺少的關鍵資訊，並從現有資料或來源的額外資訊（詮釋資料）中推斷出來。
C. 理解並將基本的資料品質原則應用於資料擷取過程的能力	對於簡單的資料品質原則如何對最終產品產生重大影響，以及如何預防後續需要額外清理的理解有限。	了解一些最基本的資料品質原則（例如：避免拼寫錯誤），但對如何將更具體的原則應用於資料擷取過程的認識有限。	了解所有基本的資料品質原則，以及如何以簡單的方式將這些原則應用於資料擷取過程。在資料擷取過程中一致地使用格式（例如：日期、國家名稱）。以簡單的方式記錄所有與資料品質相關的流程和變更。	對所有常見的資料品質原則有良好的認識，並知道如何運用這些原則來改善資料擷取過程。一致地使用資料格式，並能使用地名索引、參考資料清單或特定軟體功能來改善原始資料品質。清楚記錄所有與資料品質相關的變更和決策。

資料管理的評量標準

資料管理

技能	基礎表現 1	進步中表現 2	熟練表現 3	傑出表現 4
A. 評估生物多樣性資料集品質（即辨認問題及其類型）的能力	僅使用目視檢查來分析品質。無法區分錯誤類型。能發現必填欄位中的缺失值和嚴重的資料不一致。	只能使用非常基本的技術（例如：排序）來分析資料品質。能發現欄位名稱與內容之間的不符。能一致地辨認技術性錯誤，但只能辨認資料集中最典型的一致性錯誤。	能使用特定工具和技術來評估品質。認識一般使用和發布所需之最低程度的資料拆解/正規化。能一致地辨別技術性錯誤和資料集中大部分的一致性錯誤。	使用系統性方法分析資料集，涵蓋所有主要的資料領域。能一致地辨認資料集中的技術性和一致性錯誤。能使用其他資料來源（例如：詮釋資料或其他資料集）來辨認或推斷資料集中的一致性錯誤。
B. 執行資料格式修正的能力	只能在表格中手動進行修正。對數位資料中格式類型的使用（例如：日期、字串、數字）有基本認識。	能識別至少一種可自動修正格式錯誤的特定工具，但只能在特定情況下使用。其他情況下，使用簡單的機制（例如：「尋找與取代」）來解決問題。	能使用至少一種工具來自動修正格式錯誤。	能使用多種工具的進階功能來修正格式錯誤。

技能	基礎表現 **1**	進步中表現 **2**	熟練表現 **3**	傑出表現 **4**
C. 執行命名資料修正的能力	只能在表格中手動進行修正。僅使用個人對已知分類群的認識。	能識別至少一種可自動修正命名錯誤的特定工具，但只能在特定情況下使用。其他情況下，使用簡單的機制（例如：「尋找與取代」）來解決問題。	能使用至少一種工具來自動修正命名錯誤。能為其經常處理的分類群找到並使用適當的參考命名資訊。	能使用多種工具來修正命名錯誤。能為其專業領域以外的分類群找到並使用適當的參考命名資訊。
D. 執行地理資料修正的能力	只能在表格中手動進行修正。僅使用個人對已知地理區域的認識。	能識別至少一種可繪製地圖和/或自動修正地理資訊錯誤的特定工具，但只能在特定情況下使用。其他情況下，使用簡單的機制（例如：「尋找與取代」）來解決問題。	能使用至少一種工具來繪製地圖和/或自動修正地理資訊錯誤。能為其經常處理的區域找到並使用適當格式的參考地理資訊。	能使用多種工具來繪製地圖和/或自動修正地理資訊錯誤。能為其專業領域以外的區域找到並使用適當格式的參考地理資訊。
E. 使用特定軟體（例如：OpenRefine）作為資料清理工具的能力	能識別至少一種資料清理工具。能識別資料清理工具的主要功能（例如：OpenRefine）。	能識別多種資料清理工具。能使用資料清理軟體的一個或少數基本功能來清理資料集（例如：建立 OpenRefine 專案，使用分面、過濾、聚類、對齊）。	能使用資料清理軟體的所有基本功能來清理資料集（例如：在 OpenRefine 中使用分面、過濾、聚類、對齊）。	能使用一種或多種資料清理軟體的進階功能來清理資料集（例如：在 OpenRefine 中使用 API、正規表示式、Google Refine 表示式語言）。
F. 記錄資料轉換過程的能力	很少描述在整理、格式化或轉換資料時所做的變更。	大多時候會描述所做的變更。但描述變更的方式不一致或不完整（例如：描述變更內容，但未說明作者）。	總是記得描述所做的變更。總是一致地描述變更，使所有相同類型的編輯都能容易被辨別。	能以可重複的方式準確且一致地描述所做的變更。

資料發佈的評量標準

資料集發布**Data publishing**

技能	基礎表現 **1**	進步中表現 **2**	熟練表現 **3**	傑出表現 **4**
A. 對生物多樣性資訊（BDI）資料標準的認識	對 BDI 資料標準及 GBIF 接受哪些資料標準的認識有限或無認識。	能辨別 BDI 標準並了解 GBIF 接受哪些標準，但不知道在哪裡找到如何使用這些標準的資訊。無法辨別哪些項目是必要的。	了解 GBIF 接受的 BDI 標準。能找到已接受的核心資料集和延伸資料集清單。能依據 GBIF 對資料和詮釋資料項目的要求和/或建議標準發布資料集，並知道如何找到這些項目的定義。	了解各種 BDI 標準的特性和限制。

技能	基礎表現 **1**	進步中表現 **2**	熟練表現 **3**	傑出表現 **4**
B. 分析生物多樣性資料集是否適合透過 GBIF 發布的能力	對資料集需要符合哪些正式標準才能透過 GBIF 發布的認識有限或無認識。	了解資料集需要符合哪些正式標準才能透過 GBIF 發布，但無法評估給定的資料集是否符合這些標準。	能正確評估資料集目前是否可透過 GBIF 發布。能根據資料持有者提供的描述並在分析資料集後，為資料集指定至少一個合理的資料類型（也就是核心資料集）。	能辨別資料集的多種發布選項（在有多種發布選擇的情況下）。
C. 資料整合發布工具（IPT）使用：產生或分析高品質詮釋資料的能力	對良好詮釋資料的特性認識有限或無認識。	了解良好詮釋資料的特性，但難以識別這些特性。	了解並能夠辨別良好詮釋資料的特性。能對於現有的詮釋資料提出如何改進的建議。	了解高品質詮釋資料的特性以及如何產生。
D. 資料整合發布工具（IPT）使用：上傳/連結資料並將其對應至現有核心與延伸資料集的能力	能上傳單一檔案的資料集至 IPT，但無法成功對應至任何核心資料集。	只能上傳單一檔案的資料集至 IPT，並對應至單一類型的核心資料集，但無延伸資料集。	能將多個檔案作為單一資料集的一部分上傳至 IPT，並正確地對應至核心資料集和至少一個延伸資料集。能使用 IPT 常數 值功能。	能將多個檔案作為單一資料集的一部分上傳至 IPT，並正確地對應至核心資料集和多個延伸資料集。能使用 IPT 資料轉換功能。
E. 資料整合發布工具（IPT）使用：使用工具發布和註冊資料集的能力	能在 IPT 上檢視已發布的資料集和相關詮釋資料。能從 IPT 下載達爾文核心集檔案（DwC-A ）。能將已註冊的資料集從 IPT 導至 GBIF 入口網。	能透過上傳新的來源檔案來更新現有的已發布資料集。能重新發布檔案且無錯誤。	能成功發布和註冊新資料集。能理解並處理 IPT 中的發布錯誤訊息。	了解 IPT 資料集的版本控制。

Use Case II - 入侵種

請將自己帶入下列應用場景：

Use Case II 中有兩種不同的情境，你可以在兩者中擇一：

- 入侵種名錄
- 鱗翅目(Lepidoptera)取樣調查

在此Use Case II中的情境選擇將納入評分考量之中。

情境

追蹤入侵物種

Leucaena leucocephala (Lam.) de Wit *observed in Hawaii by Sharon Grant (licensed under CC-BY-NC 4.0)*

此段敘述是為了生物多樣性資料流通課程的練習而發展的，而接下來此練習的概念內容則是由Sharon Grant、 John Wieczorek、David Bloom 與 Laura Anne Russell發展出來的。

以下的虛構場景是基於真實資料集而來，並只允許作為教學用途使用。該原始資料集可參至Simpson A (2016). Big Island Invasive Species Committee - Pest Reports - 2005-2010. Version 4.1. United States Geological Survey. Occurrence Dataset accessed via GBIF.org on 2017-07-13.

描述

夏威夷入侵物種委員會（Hawaii Invasive Species Council，HISC）獲得一筆聯邦補助金，以與高中合作作為夏威夷州入侵物種課程的一部分。目標是增加當地對入侵物種的認識、增加資料收集，並提供報告不足區域的標註清單。專案聘請了一位全職專案經理負責監督，所有資金和分配則由 HISC 財務專員管理。

每個島嶼的入侵物種委員會（ISC）經理都獲得一筆子補助金，用來建立當地教育計劃和收集資料。這些計劃培訓高中生成為學生導師，並協助當地社區成員收集影像和資料。大島入侵物種委員會（BIISC）獲得另一筆子補助金，用來擴充其中央資料庫以容納每個 ISC 的資料、為參與學校提供專屬網站，並維護一個可供政府、公眾和學術研究使用的單一可搜尋資料入口網站。

每個島嶼選擇了兩所學校，因為這些學校位於入侵物種評估的知識和紀錄不足或缺乏的地區。老師會與當地入侵物種委員會（ISC）的推廣專員合作，製作教材，詳細說明 21 種重要的入侵植物物種，包括如何識別每個物種的生命階段以及最有效的防治方法。

夏威夷大學茂宜分校（University of Hawaii in Maui, UHM）研究所開設一門社區推廣課程。作為期末課程評量的一部分，該校四名植物學研究生正在驗證由各高中提交給當地 ISC 的影像和描述中的物種鑑定結果。

資料收集

每所高中的學生在其當地社區發起了一系列為期一天的社區調查。參與者在當地 ISC 早期偵測技術員和學生導師的指導下，前往各個地點了解照相原則，並被分配在採集活動期間要調查的路線。沿著每條路線，他們的任務是識別目標物種，並使用具有 GPS 功能的手機拍攝 1-3 張照片。

在每次社區採集活動期間，使用數位資料收集表格記錄了對 21 種目標入侵物種的每次觀察的詳細資料。參與者上傳用手機拍攝的影像，並被鼓勵使用Google 地圖點擊其位置，以記錄每次觀察的緯度和經度於表格中。表格的設計是基於 HISC 有害生物回報表格。

REPORTER INFORMATION

Report Number: T07222017172139

First Name: | Last Name:

Email: e.g. pestreport@643pest.com | Phone: e.g. 808-123-4567 or 8081234567

PEST SIGHTING INFORMATION

NOTE: Asterisk [*] and red label color indicate a required field.

*Date of Pest Sighting:

*Pest Name:

*Pest Description:
(Plant: size; flower, foliage, or fruit color / scent / orientation; habitat)
(Insect / animal: size; color; plant / host found on or nearby; habitat)

*Island of Pest Sighting:
(Please choose an island before entering information into the Location field).
--Select Island--

Location of Pest Sighting:
(Street address, cross streets, city, mile marker, place name or general area)

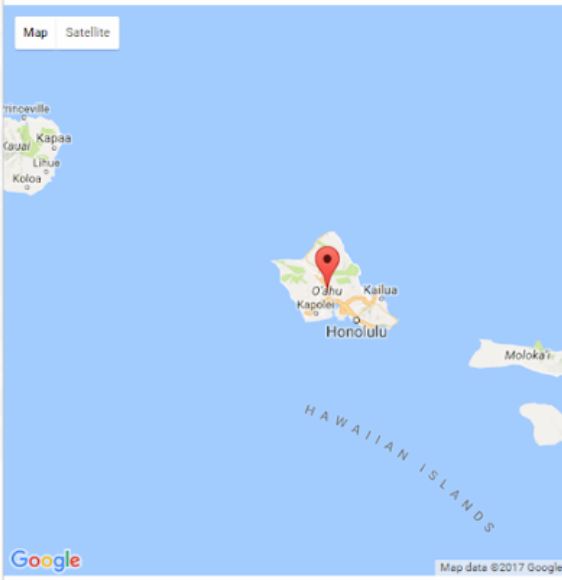
Additional Comments:

Image Upload (*Allowed File Types: .jpg, .png, .gif | Upload up to 3 images*):

Drag & drop files here ...

Browse ...

PEST SIGHTING LOCATION DETAILS



Directions:

1. The map will have automatically moved to the island chosen from the **Island** drop-down list above. If an island was not chosen or the incorrect island was chosen, please go back and make a new selection. Choosing an island is a **mandatory** step.

2. Use the **map tools** at left (Zoom In, Zoom Out, Pan) for help finding and zooming in to a desired area or location. Use the **Geo-location Search** box below to search for and zoom in to a specific geo-location (address, city, place name) in Hawaii.

3. To pinpoint the **exact location of the pest sighting**: a) click on and drag the red map marker to a particular location; or b) click on any particular location on the map to move the red map marker to that location. The **coordinates (Latitide, Longitude)** submitted with the pest report are shown in the boxes below and reflect the final position of the red map marker.

Geo-location Search:
Enter an address, city, place name...

Marker Coordinates:
Latitude | Longitude

數位資料描述

一個由夏威夷大學茂宜分校（UHM）資訊系所建立和託管的資料庫，儲存了來自線上表格的影像和資料，但這些資料不對外公開。資料以逗號分隔值（.csv）檔案格式匯出，並交給四位 UHM 研究生，讓他們使用提交的影像和描述進行分類學驗證。大島入侵物種委員會（BIISC）的地理資訊系統分析師使用 Google Maps 座標和影像 EXIF 資料來檢查觀測資料的品質，並加上任何缺失的地理參照。學生導師將所有影像檔案重新命名，使其與觀測編號相符，以便日後在 BIISC 進行交叉參照。

外來入侵物種練習表

下載 exercise sheet. (MS Word, 342 KB)

Exercise 1

規劃階段

你是當地入侵物種委員會（ISC）經理，由於課程計畫和社區調查的成功，你所在島嶼上另外 10 所學校希望在下一年建立自己的專案。你希望能配合他們，但你的 ISC 經費將在今年年底到期。夏威夷入侵物種委員會（HISC）表示他們會積極考慮下一年度擴展你的計畫的小額補助金申請，而大島入侵物種委員會（BIISC）也提供支援。

Exercise 1a

分析**增**加學校數量的財務影響

1. 評估以下選項以擴大參與學校的數量。您只能選擇其中的兩個選項，所以需要明智地選擇。
2. 請使用練習表格提供您的答案。

選項

1. 聘請額外的暑期實習生在當地 ISC工作，協助協調問卷調查。
2. 提供財務支援給 BIISC 為每所新學校建立網站。
3. 提供研究生薪資報酬。您無法為全部四位研究生支付相當於正常薪資的費用，但可以負擔其中兩位的兼職職位費用。
4. 和軟體公司簽約，來建置能直接從線上表單自動提取資料的資料庫。該系統將同時包含用來操作資料、並可以csv格式輸出的管理介面。
5. 資助四場公眾擴大行動(如：生物閃電戰)，藉以提升社群關注、增進志工參與。
6. 向學校教師提供可重複應用的訓練課程教案，並教導他們該如何準備向BIISC繳交的資料

Exercise 1b

指派角色

此新專案中，有以下幾位人員可以用來進行資料處理、數位化

1. 請指派角色與任務，來最大化資料處理與轉換的效率，以此來盡可能地產生最高品質的資料
2. 請使用練習表格提供您的答案。

角色

- BIISC GIS 分析師：精通電腦使用、地理資訊系統（GIS）和資料分析工具

- ISC 經理：良好的電腦使用能力
- ISC 外展助理：良好的野外物種鑑定能力；基本電腦使用能力。社群媒體專家。
- 學生導師 1：基礎分類學知識。基本電腦使用能力。
- 學生導師 2：基礎分類學知識。基本電腦使用能力。
- 植物學學生 1：進階的分類學知識。程式設計能力。
- 植物學學生 2：進階的分類學知識。
- 植物學學生 3：進階的分類學知識。
- 植物學學生 4：進階的分類學知識。

Exercise 2

資料獲取

BIISC 正在計畫透過發布資料集至 GBIF 公開專案的所有資料。身為 BIISC 的外展助理，你必須判斷相關的達爾文核心標準（Darwin Core）的欄位以提供線上表單資料的對照。你注意到進行驗證的研究生已在資料表單中加入描述物種和地點的其他資料。為了能夠使這些資料對照達爾文核心標準，你需要擴展資料結構以彙整來自線上表單的資料以及新增的分類學和地理參考資料。

1. 下載 UC2-IS-2-ForCapture.csv. (7.1 MB)
2. 使用下載的資料集，製作一個試算表作為擴展資料結構的範例，並列出你認為與 達爾文核心標準相關的欄位。
3. 使用練習表提供你的答案並繳交試算表。

Exercise 3

資料管理

暑假期間，HISC 總部的實習生從線上表單的原始出現紀錄收集並擴增來建立了物種名錄。身為HSC專案負責人，你現在必須在資料發布前執行最後的品質檢查。

1. 下載 UC2-IS-3-ForCleaning.xlsx. (156 KB)
2. 評估資料集並判斷錯誤的類型。
3. 找出可能的修正方法並盡可能修正這些錯誤。
4. 使用練習表提供你的答案並繳交試算表。

Exercise 4

資料集發布**Data publishing**

HISC 現在準備將物種名錄和相關出現紀錄發布到 GBIF。在這個練習中，你將擔任專案負責人。你的責任是：透過 GBIF 網路發布已清理的清單資料和相關出現紀錄。

1. 下載 UC2-IS-4-ForPublication.xlsx. (99 KB)
2. 使用先前提供的 IPT 安裝來發布這個資料集。
3. 使用練習表提供你的答案並連結到已發布的資料集。

案例二 - 鱗翅目目擊紀錄

請將自己帶入下列應用場景：

Use Case II 中有兩種不同的情境，你可以在兩者中擇一：

- 入侵種名錄
- 鱗翅目(Lepidoptera)取樣調查

在此Use Case II中的情境選擇將納入評分考量之中。

情境

跨國鱗翅目採樣

Papilio machaon Linnaeus, 1758 *observed in Israel by רניו רמוע (licensed under* CC-BY-NC 4.0*)*

此敘述是為生物多樣性資料流通課程的實作練習而開發，練習概念和內容由 Alberto González-Talaván 基於 Alberto González-Talaván、Danny Vélez、Larissa Smirnova、Laura Russell、Mélianie Raymond 和 Nicolas Noé 的先前工作發展而成。這是一個虛構的場景並僅供教學使用。

描述

國際蝴蝶愛好者協會（ International Butterfly Amateur Network，IBAN）自2009年起提供各國業餘觀察團體來記錄蝴蝶（鱗翅目）出現資料的框架。廣大的業餘觀察者網絡使用基於 Pollard walks 的標準協議來收集這些資訊，並將紙本記錄寄送至各國辦事處。部分辦事處將這些資訊數位化為試算表，但其他辦事處因缺乏人力而將紙本記錄寄送至 IBAN 進行處理。IBAN 根據這些國家會員提供的目

擊資料製作年度報告，包含更新分布的地圖和部分重要物種的族群趨勢分析。

IBAN 總部主要由志工組成。隨著公民科學日益普及，以及大眾對蝴蝶這種迷人生物的興趣與日俱增，每年收到的資料量持續增加，紙本資料表在尚未數位化的情況下快速堆積。IBAN 指導委員會正試圖找出一個更有效率且靈活的工作流程來建立數位資料，因為他們希望開始定期在線上發布這些資料。他們也想要開始處理志工們已經用手機和平板等行動裝置拍攝的數位照片。他們最終目標是提升網路知名度，並加強與地方及區域政府的合作，以影響相關國家的鱗翅目保育政策。

目前 IBAN 與收集資料的業餘愛好者之間並沒有正式協議來規範資料的使用方式。指導委員會擔心，當他們開始在線上發布資料時，會需要將相關規範正式化。

資料收集

建議的調查方法——穿越線調查法（Pollard walks）——是基於 300 到 600 公尺長的穿越線，每 50 公尺被分為一個小段，而每條穿越線只涵蓋單一棲地類型。

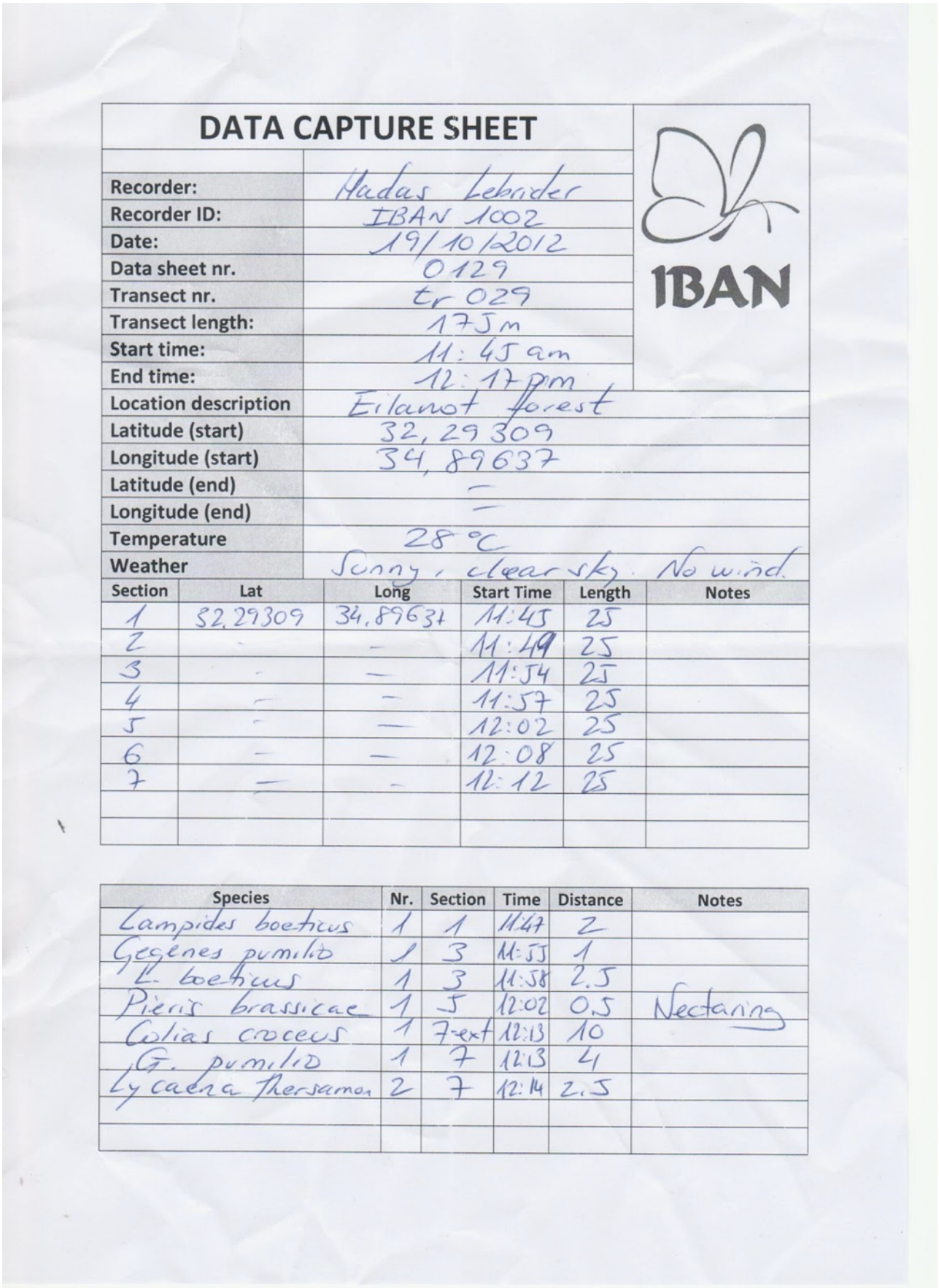
在每次調查中，穿越線調查員需要計算在穿越線兩側 5 公尺範圍內所有可見的鱗翅目物種，其中特殊行為（如：產卵或吸蜜）以及發育階段（如：幼蟲或卵）也都需要記錄。

對大多數國家而言，這些調查活動從每年十月初到次年六月底，每兩週進行一次。

目前相關的品質管理機制已經建立完成：每筆回報的出現紀錄都會被標記為「待核准」。在特定的分類學專家驗證後，紀錄狀態才會改為「已核准」。在非正常季節或分布區域發現的物種會被標記進行額外驗證。

穿越線調查開始時會記錄當天的時間和天氣狀況，並沿著穿越線記錄看到的每個物種的個體數量。未能辨識的物種則依照科別或預先定義的二到三個相似物種組合來計數並記錄；在 5 公尺範圍外看到的蝴蝶則記錄為「額外 + 最近區段編號」（例如：5-extra）。穿越線調查結束時間也會被記錄。

類比（紙本）資料收集範例：



DATA CAPTURE SHEET

IBAN

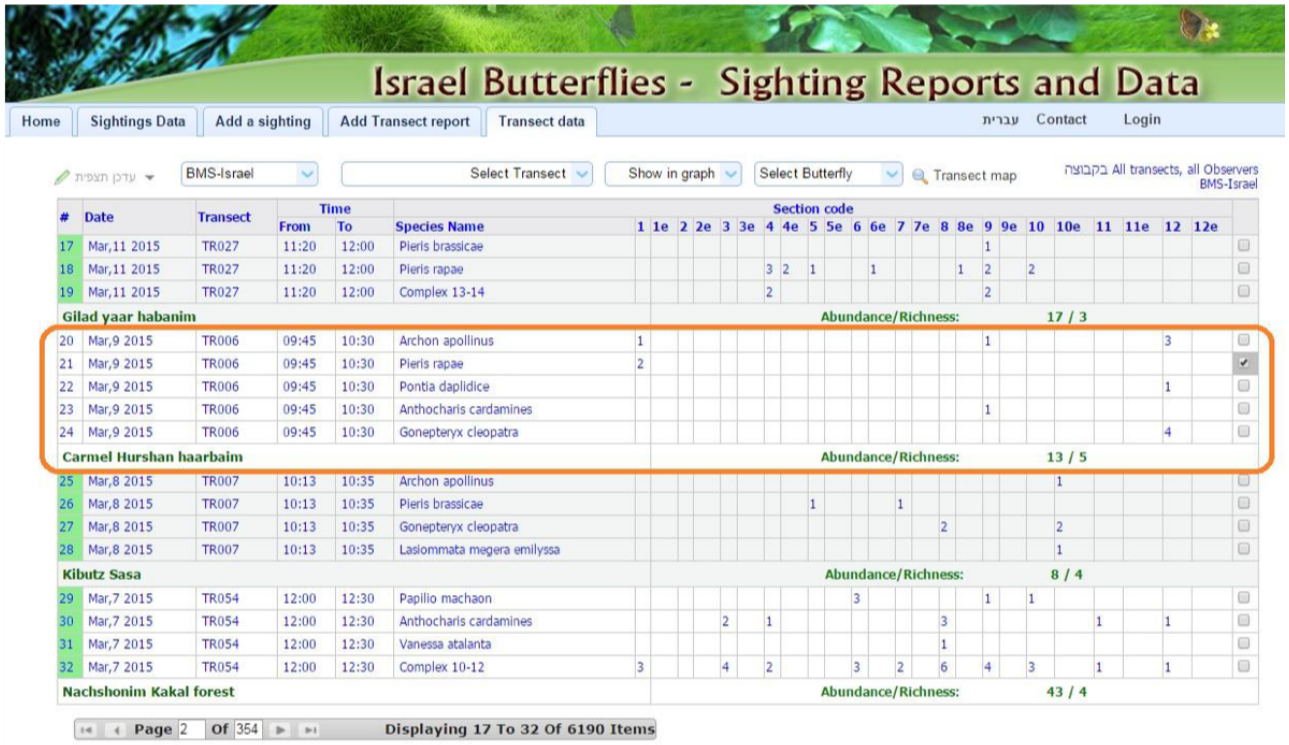
Recorder:	Hadas Lebiush
Recorder ID:	IBAN 1002
Date:	19/10/2012
Data sheet nr.	0129
Transect nr.	tr 029
Transect length:	175 m
Start time:	11:45 am
End time:	12:17 pm
Location description	Eilanot forest
Latitude (start)	32,29309
Longitude (start)	34,89637
Latitude (end)	–
Longitude (end)	–
Temperature	28 °C
Weather	Sunny, clear sky. No wind.

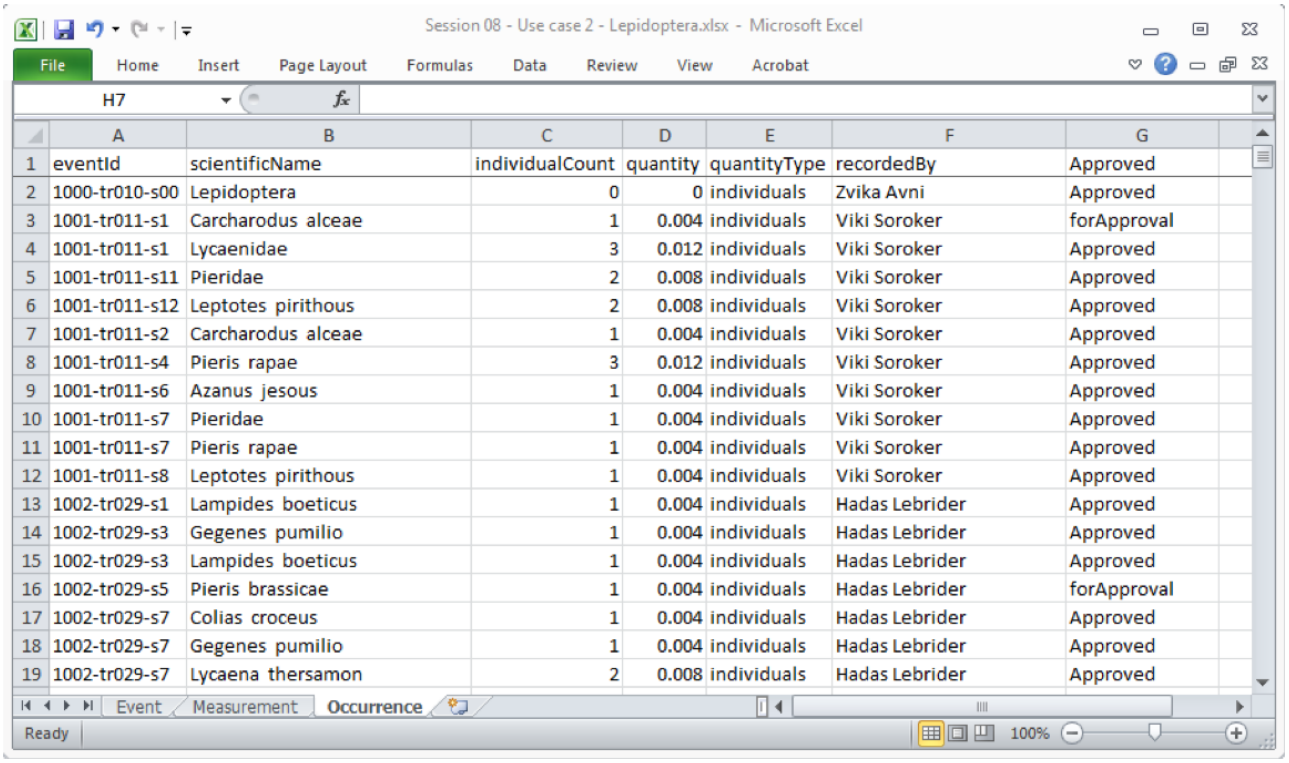
Section	Lat	Long	Start Time	Length	Notes
1	32,29309	34,89637	11:45	25	
2	–	–	11:49	25	
3	–	–	11:54	25	
4	–	–	11:57	25	
5	–	–	12:02	25	
6	–	–	12:08	25	
7	–	–	12:12	25	

Species	Nr.	Section	Time	Distance	Notes
Lampides boeticus	1	1	11:47	2	
Gegenes pumilio	1	3	11:55	1	
L. boeticus	1	3	11:58	2.5	
Pieris brassicae	1	5	12:02	0.5	Nectaring
Colias croceus	1	7-ext	12:13	10	
G. pumilio	1	7	12:13	4	
Lycaena thersamon	2	7	12:14	2.5	

數位資料描述

部分國家辦公室使用志工團隊來數位化紙本紀錄並製作數位試算表。這些試算表非常簡單，共包含三個資料表：第一個記錄與調查活動相關的資訊，第二個記錄天氣狀況，第三個記錄業餘愛好者觀察到的物種與個體數量。





鱗翅目調查表

下載 exercise sheet. (MS Word, 342 KB)

Exercise 1

規劃階段

類比資料（紙本記錄）抵達 IBAN
總部的數量很快就會超過他們的數位化處理能力，因此指導委員會決定重新考慮這個在過去幾年中未經管理而持續成長的工作領域的現行方法。到目前為止，工作是這樣組織的：

- 紙本記錄透過郵寄抵達，秘書開啟包裹並整理這些記錄。
- 有五位具備基本電腦技能的志工使用兩台共用電腦來數位化紙本記錄。這些志工本身也是公民科學家，所以他們熟悉鱗翅目的分類學，也了解 IBAN 總部所在國家出現的物種。
- 數位化工作人員依照自己的時間進出，所以他們通常會透過電話確認電腦是否可用。有時會發生時間衝突，有些人因為兩台電腦都在使用中而必須回家，有時兩台電腦則無人使用。
- 在進行資料數位化時，他們通常從紙本堆中一次挑選一份來處理（如果他們能夠處理的話）。常見的問題包括：
  - 數位化人員不認識觀察到的物種（會發生拼寫錯誤），
  - 數位化人員不熟悉進行調查活動的地區，
  - 數位化人員無法辨認手寫字跡或某些評論所使用的語言。
- 一位分類學專家負責處理所有數位化的表格，並根據這些資料製作報告和分布圖。通常她需要捨棄約 15% 的數位化資料，因為這些資料存在不一致、拼寫錯誤或其他她沒有時間檢查的錯誤。

Exercise 1a

分析他們新數位化計畫的財務部分

指導委員會正在分析以下新數位化計畫的選項，這些選項都會影響他們已經縮減的預算。他們知道只能實施其中「兩個」選項，所以需要明智地選擇。請使用練習表為他們提供建議，說明應該選擇哪兩個選項以及選擇的原因。

1. 選項 1：再購買三台電腦，讓所有數位化人員能同時工作。
2. 選項 2：為國家辦公室提供財務支持、購買文件掃描器，並以電子方式傳送／分享記錄取代郵寄。
3. 選項 3：提供數位化人員財務報酬。雖然無法支付所有五位志工相當於正職的薪資，但可以支付三位志工兼職職位的費用。
4. 選項 4：購買現有的英文版生物多樣性數位化軟體，其具備分類學輸入檢查功能和內建的地理資訊更正輔助工具。
5. 選項 5：聘請軟體開發公司開發客製化的數位化軟體。以與商業軟體相同的價格，開發人員將提供使用當地語言的解決方案，完全符合原始資料架構，並提供網頁資料入口網站來展示數位化成果。
6. 選項 6：為五位數位化人員組織一個課程，提升他們在分類學、電腦使用和生物多樣性資訊學標準方面的技能。

Exercise 1b

指派角色

這些是數位化工作可用的人力資源，你會如何分配角色以最佳化數位化過程的效率並產出最高品質的資料？請使用練習表提供您的答案。

1. 行政助理：無分類學知識。基本電腦使用能力。可閱讀 3 種語言。
2. 志工 1：基礎分類學知識。基本電腦使用能力。
3. 志工 2：基礎分類學知識。基本電腦使用能力。
4. 志工 3：基礎分類學知識。基本電腦使用能力。可閱讀 3 種語言。
5. 志工 4：基礎分類學知識。基本電腦使用能力。可閱讀 3 種語言。
6. 志工 5：基礎分類學知識。進階電腦使用能力（包含 GIS 和資料分析工具）。

7. 分類學專家：進階分類學知識。進階電腦使用能力（包含 GIS 和資料分析工具）。

Exercise 2

資料獲取

假設你是 IBAN 總部接收到紙本紀錄的數位化志工之一。你收到了兩份紙本紀錄。

1. 下載紀錄 1 和 2 UC2-LS-2-ForCapture.zip. (943 KB)
2. 你會使用什麼資料結構來呈現這些紀錄中的資料？
3. 請使用此結構和紀錄中的資料建立試算表。
4. Use the exercise sheet to provide your answers and submit the spreadsheet created in the previous step.

Exercise 3

資料管理

假設你是具有進階電腦技能的志工之一，負責資料品質管理問題。你的主要任務是減少目前因錯誤和不一致而在處理前被丟棄的資料量（約 15%）。你收到了一個數位化工作的原始資料集。

1. 下載 UC2-LS-3-ForCleaning.xlsx. (44 KB)
2. 評估資料集並判斷錯誤的類型。
3. 找出可能的修正方式，並盡可能修正其中的錯誤。
4. 使用練習表提供你的答案並繳交試算表。

Exercise 4

資料集發布**Data publishing**

在這個練習中，你將扮演與 IBAN 總部合作的分類學專家。你先前的責任（撰寫年度報告和製作基礎分布地圖）已經交給志工，而你現在有了新的責任：透過 GBIF（全球生物多樣性資訊機構）網絡在線上發布已清理的資料。負責資料品質管理的志工已提供了要發布的資料集。

1. 下載 UC2-LS-4-ForPublication.xlsx. (58 KB)
2. 使用先前提供的 IPT 安裝來發布這個資料集。
3. 使用練習表提供你的答案並連結到已發布的資料集。

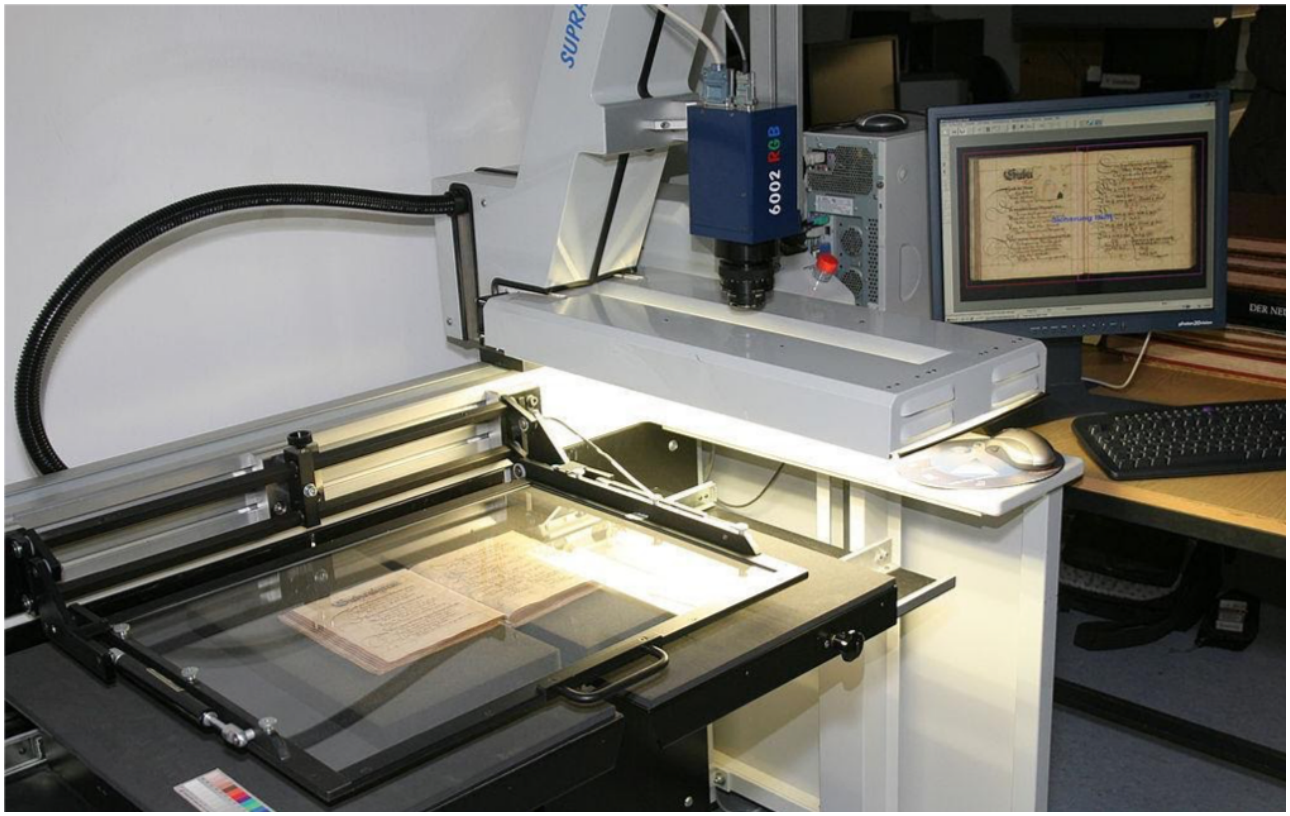
案例三 - 文獻中的鳥類

請將自己帶入下列應用場景：

案例三將被評分

情境

來自文獻「**Danish Lighthouses** 落下的鳥類，**1883-1939**（**Birds fallen at Danish Lighthouses, 1883–1939**）」的資料流通專案



https://upload.wikimedia.org/wikipedia/commons/b/bc/Wikimedia_Rechenbuchdigitalisierung_2006-05-24_01.JPG *[高解析度掃描器]* 由 *Heiko Hornig*
執行的書籍數位化專案（採用*https://creativecommons.org/licenses/by-sa/2.5/deed.en* 創用 *CC* 授權條款 *CC BY-SA 2.5* 授權）

這個敘述是作為生物多樣性資料流通課程實作練習的基礎而開發的，練習概念和內容由Alberto González-Talaván、Andrea Hahn、Laura Russell 和 Sharon Grant 開發。這是基於Alberto González-Talaván、Danny Vélez、Larissa Smirnova、Laura Russell、Mélianie Raymond 和 Nicolas Noé 先前改編的版本。

這是基於真實專案和資料集的虛構情境，僅供教學用途。原始專案（https://danbif.dk/se-eksempler/fyrfaldne-fugle/）和原始資料集（https://www.gbif.org/dataset/ad331dcc-d0fa-4816-b1e6-d36f9f899c49）歸屬於丹麥 GBIF（https://danbif.dk/）。

描述

丹麥自然歷史博物館（NHM-DK）是哥本哈根大學的研究中心。他們的圖書館是國家圖書館協會的成員，最近獲得了國家資金，以利將其成員持有的資源放上網路。NHM-DK 想要開始將他們圖書館中的野外筆記本、期刊出版物和書籍數位化，其中一些具有重要的歷史價值。

在與合作夥伴進行簡短諮詢後，NHM-DK 收到了來自 Nordjylland 國家公園管理辦公室主任的建議。他們希望將一份特定的經典文獻彙編數位化，用於他們正在進行的專案：「Danish Lighthouses 的鳥類，1883-1939」（丹麥文為「Fuglene ved de danske Fyr, 1883-1939」）。他們想要在實體展覽專案中使用這些書籍中任何有關兩座燈塔（Lodbjerg Fyr 和 Hanstholm Fyr）的出現紀錄資料。

NHM-DK 已經開始與他們國家的 GBIF（全球生物多樣性資訊機構）——DanBIF 討論這些卷冊中資訊的流通問題，主要是為了保存其內容以供未來使用，並為所有人提供線上存取。在 DanBIF 的參與下，計劃將提取的資料發布並註冊到 GBIF。由於 GBIF 要求所有發布的資料都必須附加授權條款，博物館已決定以允許註明出處使用資料的創用 CC 授權條款（CC-BY）發布資料。

專案需要的 IT 服務由哥本哈根大學的技術部門提供，如同所有博物館的數位專案一樣。

NHM-DK 副館長正在協調這項工作，並建立了工作的整體框架：

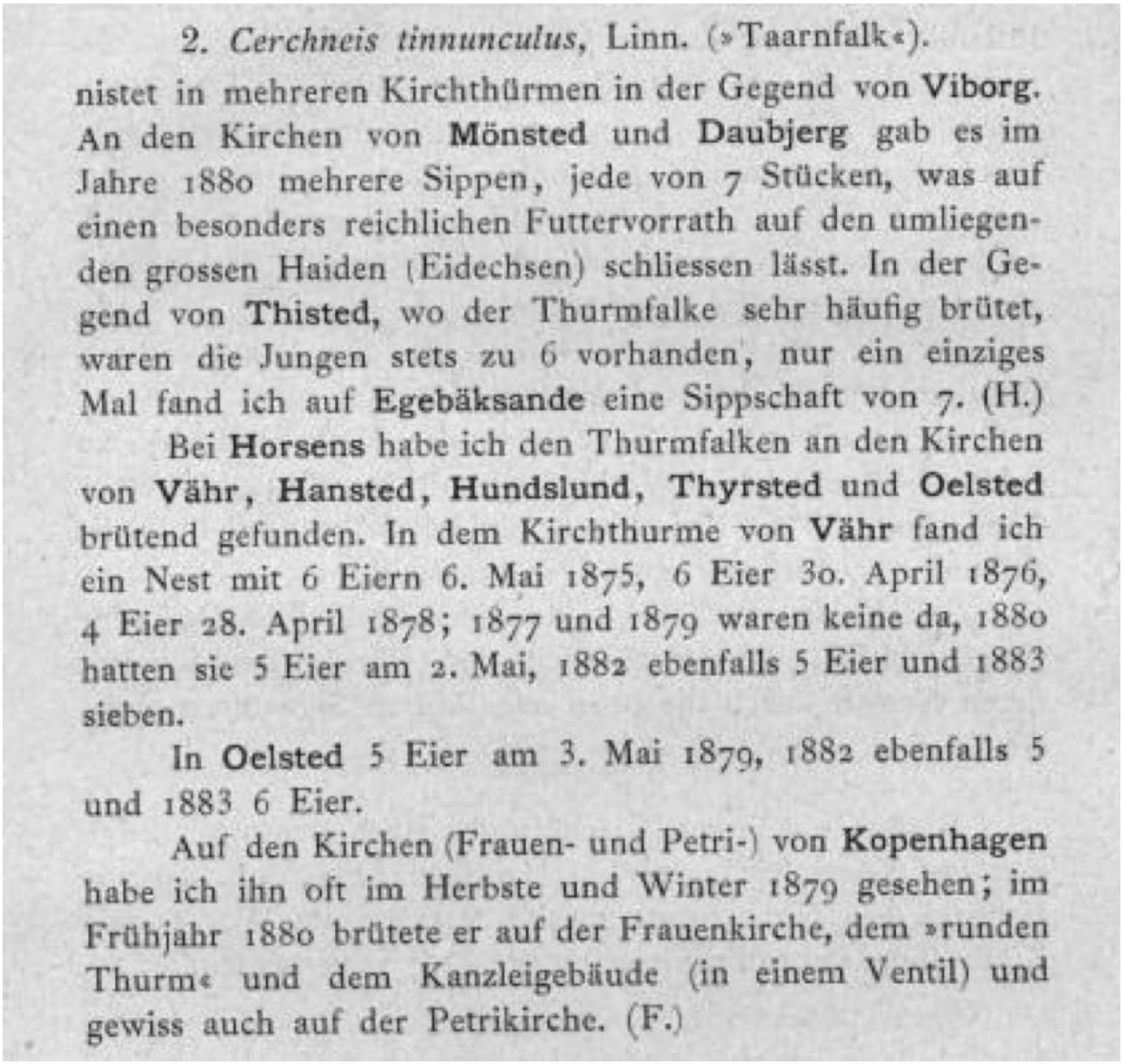
1. 博物館將由兩名受過圖書館掃描器使用培訓的圖書館工作人員進行文獻數位化，他們會小心處理這些脆弱的卷冊。他們還會透過 OCR（光學字元識別）軟體從掃描檔中提取文字。
2. 來自哥本哈根鳥類學會（COS）的三名志工經常與博物館合作，並熟悉該地區的鳥類，他們將協助並完成將掃描的 PDF 資料轉換為試算表格式的工作。他們需要到博物館使用圖書館的電腦才能存取儲存在博物館內部網路（私人網路）中的檔案。
3. NHM-DK 鳥類部門的鳥類學研究員將帶領團隊共同負責分類檢查、資料典藏、清理、格式和轉換，並監督已發布資料集的詮釋資料輸入。團隊包括一名來自瑞典的合作研究員和兩名博士後研究生。他們被選定負責這項任務是因為他們習慣處理數位生物多樣性資料。他們都將使用自己的工作電腦。
4. DanBIF 節點管理員將確保該機構在 GBIF 正確註冊為資料發布者，並確保副館長和鳥類學研究員擁有合適的憑證和存取權限，以使用 DanBIF 的資料整合發布工具（IPT）上傳和發布資料。

原始資料收集

在 1883 年至 1939 年期間，丹麥共有 45 座運作中的燈塔和燈船。這些燈塔在 1886 年至 1939 年鳥類遷徙期間被多種鳥類使用。這些鳥類的出現和活動主要由燈塔管理員記錄，他們還收集標本並將其送往哥本哈根的博物館。這些鳥類由博物館的典藏經理仔細保存和編目，至今仍保存在那裡。管理員還記錄了觀測鳥類期間的夜間天氣狀況。

類比資料（紙本資料）描述

這是從一本書中記錄的一系列物種觀測資料的描述範例（書籍使用德文，只有物種的俗名使用丹麥語）。



2. *Cerchneis tinnunculus*, Linn. (»Taarnfalk«).

nistet in mehreren Kirchthürmen in der Gegend von **Viborg**. An den Kirchen von **Mönsted** und **Daubjerg** gab es im Jahre 1880 mehrere Sippen, jede von 7 Stücken, was auf einen besonders reichlichen Futtervorrath auf den umliegenden grossen Haiden (Eidechsen) schliessen lässt. In der Gegend von **Thisted**, wo der Thurmfalke sehr häufig brütet, waren die Jungen stets zu 6 vorhanden, nur ein einziges Mal fand ich auf **Egebäksande** eine Sippschaft von 7. (H.)

Bei **Horsens** habe ich den Thurmfalken an den Kirchen von **Vähr**, **Hansted**, **Hundslund**, **Thyrsted** und **Oelsted** brütend gefunden. In dem Kirchthurme von **Vähr** fand ich ein Nest mit 6 Eiern 6. Mai 1875, 6 Eier 30. April 1876, 4 Eier 28. April 1878; 1877 und 1879 waren keine da, 1880 hatten sie 5 Eier am 2. Mai, 1882 ebenfalls 5 Eier und 1883 sieben.

In **Oelsted** 5 Eier am 3. Mai 1879, 1882 ebenfalls 5 und 1883 6 Eier.

Auf den Kirchen (Frauen- und Petri-) von **Kopenhagen** habe ich ihn oft im Herbste und Winter 1879 gesehen; im Frühjahr 1880 brütete er auf der Frauenkirche, dem »runden Thurm« und dem Kanzleigebäude (in einem Ventil) und gewiss auch auf der Petrikirche. (F.)

掃描與翻譯資料描述

這是上述類比資料（紙本資料）的掃描和翻譯範例。

Output of the OCR software	Translation into English
2. *Cerchneis tinnunculus*, Linn. (»Taarnfalk«) nistet in mehreren Kirchthürmen in der Gegend von **Viborg**. An den Kirchen von **Mönsted** und **Daubjerg** gab es im Jahre 1880 mehrere Sippen, jede von 7 Stücken, was auf einen besonders reichlichen Futtervorrath auf den umliegenden grossen Haiden (Eidechsen) schliessen lässt. In der Gegend von **Thisted**, wo der Thurmfalke sehr häufig brütet, waren die Jungen stets zu 6 vorhanden, nur ein einziges Mal fand ich auf **Egebäksande** eine Sippschaft von 7. (H.) Bei **Horsens** habe ich den Thurmfalken an den Kirchen von **Vähr**, **Hansted**, **Hundslund**, **Thyrsted** und **Oelsted** brütend gefunden. In dem Kirchthurme von **Vähr** fand ich ein Nest mit 6 Eiern 6. Mai 1875, 6 Eier 30. April 1876, 4 Eier 28. April 1878; 1877 und 1879 waren keine da, 1880 hatten sie 5 Eier am 2. Mai, 1882 ebenfalls 5 Eier und 1883 sieben. In **Oelsted** 5 Eier am 3. Mai 1879, 1882 ebenfalls 5 und 1883 6 Eier. Auf den Kirchen (Frauen- und Petri-) von **Kopenhagen** habe ich ihn oft im Herbste und Winter 1879 gesehen; im Frühjahr 1880 brütete er auf der Frauenkirche, dem »runden Thurm« und dem Kanzleigebäude (in einem Ventil) und gewiss auch auf der Petrikirche. (F.)	2. *Cerchneis tinnunculus*, Linn. ("Taarnfalk") nests in several steeples around **Viborg**. At the churches of **Mönsted** and **Daubjerg** there were several family groups in 1880, each of 7 individuals, suggesting a particularly abundant source of food on the surrounding heather (lizards). In the area of **Thisted**, where the tower falcon broods very often, young were always present in broods of 6, only once did I find a group of 7 on **Egebäksande** (H.) In **Horsens** I found kestrels brooding on the churches of **Vähr**, **Hansted**, **Hundslund**, **Thyrsted** and **Oelsted**. In the steeple of **Vähr** I found a nest with 6 eggs on 6 May 1875, 6 eggs on 30 April 1876, 4 eggs on 28 April 1878; in 1877 and 1879 there were none, on 2 May 1880 they had 5 eggs, in 1882 also 5 eggs, and in 1883 seven. In **Oelsted** 5 eggs on 3 May 1879, in 1882 also 5, and in 1883 6 eggs. On the churches (Our Lady's and St. Peter's) of **Copenhagen**, I have often seen it in the autumn and winter of 1879; in spring 1880, it brooded on Our Lady's church, the "round tower" and the law firm building (in a valve) and certainly also on St. Peter's Church. (F.)

數位資料描述

研究這本書的摘錄，哥本哈根鳥類學會的志工建議從掃描和翻譯的文本中提取以下資料：

- 書中出現的科學名稱
- 書中出現的丹麥語俗名
- 地點
- 年/月/日
- 觀測個體數量
- 性別
- 生命階段
- 備註
- 出現紀錄的數位化書籍頁面的 URL

文獻中的鳥類工作表

下載 exercise sheet. (MS Word, 342 KB)

Exercise 1

規劃階段

這個團隊需要開發一個數位化文獻資源的永續工作流程，摘要紙本文獻中有關生物多樣性的資訊，並透過 GBIF 發表在網路上。他們需要制定一個在國家圖書館協會資助結束後仍能永續的計劃。

《scenario-4，情境》部分包含副主任構想工作流程的概要描述。請根

據工作流程和附帶文字完成以下任務：

1. 辨識參與此專案的不同利益關係者
2. 確定他們的隸屬關係，並將每個人分配到相應的利益關係者群組
3. 辨識與他們相關的角色，並分配他們目前負責的任務
4. 對工作流程進行批判分析，辨識潛在風險和缺口，並提出改善工作流程的建議，藉此最佳化數位化專案的效率並產出最高品質的資料。
5. 請使用練習表格提供您的答案。

Exercise 2

資料獲取

書籍的掃描和文字識別（OCR）已經完成。現在必須從這些來源提取出現紀錄資料，並編譯成試算表格式。

由於原始資料是德文，為了讓資料在線上發布時適用範圍更廣，專案經理希望將其轉換為英文版本。

> 2. Cerchneis tinnunculus, Linn. ("Taarnfalk") nests in several steeples around Viborg.
>
> At the churches of Mönsted and Daubjerg there were several family groups in 1880, each of 7 individuals, suggesting a particularly abundant source of food on the surrounding heather (lizards). In the area of Thisted, where the tower falcon broods very often, young were always present in broods of 6, only once did I find a group of 7 on Egebäksande (H.)
>
> In Horsens I found kestrels brooding on the churches of Vähr, Hansted, Hundslund, Thyrsted and Oelsted. In the steeple of Vähr I found a nest with 6 eggs on 6 May 1875, 6 eggs on 30 April 1876, 4 eggs on 28 April 1878; in 1877 and 1879 there were none, on 2 May 1880 they had 5 eggs, in 1882 also 5 eggs, and in 1883 seven.
>
> In Oelsted 5 eggs on 3 May 1879, in 1882 also 5, and in 1883 6 eggs.
>
> On the churches (Our Lady's and St. Peter's) of Copenhagen, I have often seen it in the autumn and winter of 1879; in spring 1880, it brooded on Our Lady's church, the "round tower" and the law firm building (in a valve) and certainly also on St. Peter's Church. (F.)

1. 請扮演一位負責將翻譯文本轉換為個別出現紀錄的志工，這些出現紀錄需要分配唯一編號。
2. 請使用上述範例中由 Chr. Fr. Lütken 記錄的資料，並根據 《birds-digital-data-description》中列出的資料欄位建立試算表。
3. Use the exercise sheet to provide your answers and submit the spreadsheet created in the previous step.

在使用的範例中，個別的出現紀錄並不總是包含完整的試算表欄位資料。

Exercise 3

資料管理

資料現在已經由哥本哈根鳥類學會的志工整理成試算表格式。作為鳥類部門的典藏總監，你被指派負責該資料集的資料品質問題。

透過追溯地理參照（retrospective georeferencing），座標已與地點一起添加到資料集中，但

沒有其他更高層級的地理資訊。由於所有觀察都是在丹麥進行，因此可以輕易增加洲別和國家資訊。此外，目前資料集裡僅提供科學名稱。你可以使用 OpenRefine 等軟體工具來推導更高層級的分類系統。你也注意到數位化人員發生了一些拼字錯誤。

1. 下載 UC3-DL-3-ForCleaning.zip. (45 KB)
2. 辨識並更正任何無效的年份。
3. 檢核並更正分類系統。
4. 檢核兩個給定地點的座標是否正確。如有錯誤請更正。座標應該使用十進位格式。
5. 使用現有資料增加任何可推導的缺失元素
6. 請記得保留原始提供的資訊，並將你的更改和假設紀錄個別記錄為詮釋資料的一部分。
7. Use the exercise sheet to provide your answers and submit the cleaned text file extracted from the step 1.

資料集應僅包含 1883-1939 年間的資料

Exercise 4

資料集發布**Data publishing**

在這個練習中，你將扮演負責透過 GBIF（全球生物多樣性資訊機構）網路線上發布已清理資料的工作人員。你已獲得多媒體檔案和歷史辨識檔案，這些檔案要與觀察資料一起發布。負責資料品質的工作人員已為你提供已清理的資料集以供發布。

1. 下載 UC3-DL-4-ForPublication.zip. (65 KB)
2. 使用先前提供的 IPT 安裝來發布這個資料集。
3. 使用練習表提供你的答案並連結到已發布的資料集。

最終作業

你需要完成並提交使用案例 II 和使用案例 III 作為評估依據。

案例二

有兩個選項（《use-case-ii-invasive-species，入侵物種》或《use-case-ii-lepidoptera-sightings，鱗翅目目擊》）。你只需要選擇其中之一完成作業。

需要繳交的檔案：

- 完成的練習表（接受 MS Word Doc 或類似格式）
- 資料擷取試算表（接受 MS Excel、csv、txt 或類似格式）
- 清理／標準化的資料集（接受 MS Excel、csv、txt 或類似格式）

案例三

只有一個選項（《use-case-iii-birds-from-literature，文獻中的鳥類》）。

需要繳交的檔案：

- 完成的練習表（接受 MS Word Doc 或類似格式）

- 資料擷取試算表（接受 MS Excel、csv、txt 或類似格式）
- 清理／標準化的資料集（接受 MS Excel、csv、txt 或類似格式）

所有繳交的檔案都需要包含使用案例和練習編號以及你的姓名。例如：Russell-UC2-IS-exercise-sheet.docx、Russell-UC2-IS-2.xlsx、Russell-UC2-IS-3.xlsx。所有檔案必須以英文繳交。如有任何問題，請聯繫 training@gbif.org。

作業繳交

可以從課程的線上（HTML）版本繳交作業。

課程評估

完成課程評估

重要文件

以下參考資料提供本課程相關主題的更多細節，所有連結都會在新視窗／分頁中開啟。

達爾文核心集（DwC）

- Darwin Core Terms: A quick reference guide
- Simple DarwinCore
- Darwin Core Questions & Answers
- Darwin Core extensions registered with GBIF

資料發布

- Quick guide to publishing data through GBIF
- How to publish biodiversity data through GBIF.org
- Become a data publisher with GBIF
- Best Practices for Publishing Biodiversity Data from Environmental Impact Assessments
  GBIF Secretariat & IAIA: International Association for Impact Assessment (2020).
- Guidance for private companies to become data publishers through GBIF: Template document to support the internal authorization process to become a GBIF publisher
  Rui Figueira, Pedro Beja, Cristina Villaverde, Miguel Vega, Katia Cezón, Tainan Messina, Anne-Sophie Archambeau, Rukaya Johaadien, Dag Endresen & Dairo Escobar (2020).
- Publishing DNA-derived data through biodiversity data platforms
  Anders F. Andersson, Andrew Bissett, Anders G. Finstad, Frode Fossøy, Marie Grosjean, Michael Hope, Thomas S. Jeppesen, Urmas Kõljalg, Daniel Lundin, R. Henrik Nilsson, Maria Prager, Cecilie Svenningsen & Dmitry Schigel (2020).
- Classes of datasets supported by GBIF
- GBIF data quality requirements for publishing
- GBIF data licenses
- Checklist core templates

- Occurrence core templates
- Sampling event core templates
- Sampling event data best practices
- Sharing images, sounds and videos on GBIF
- Data papers
- Published data papers

資料發布：資料整合發布工具（IPT）

- The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet
  Robertson et al. (2014)
- To install IPT or not to install IPT
- IPT data hosting centres
- IPT Install / Set up webinar
- Installing the IPT video
- IPT in Practice demonstration video

數位化

- iDigBio Digitization Resources
- iDigBio Collections Digitization Workflows
- iDigBio Digitization Workflows and Protocols
- iDigBio specimen image capture guide
- Canadensys 10-step guide to managing images with your biodiversity data

GBIF

- What is GBIF
- Strategic Plan
- Become a member
- Science Review
- Establishing an Effective GBIF Participant Node: Concepts and general considerations
  GBIF Secretariat (2019).

地理參照座標

- Georeferencing Best Practices
  Arthur D. Chapman & John R. Wieczorek (2020).
- Georeferencing Quick Reference Guide
  Paula F. Zermoglio, Arthur D. Chapman, John R. Wieczorek, Maria Celeste Luna & David A. Bloom (2020).
- Georeferencing Calculator Manual
  David A. Bloom, John R. Wieczorek & Paula F. Zermoglio (2020).

- Georeferencing resources

入侵物種

- GRISS - Global Register of Introduced and Invasive Species
- TrIAS - Tracking Invasive Alien Species

生命地圖集

- Living Atlases
- ALA key technical documentation

雜項

- VertNet Guide to opening text files in Excel
- VertNet data licensing guide

OpenRefine

- OpenRefine documentation
- OpenRefine regular expressions
- Guía para la limpieza de datos sobre biodiversidad con OpenRefine
  Paula F. Zermoglio, Camila A. Plata Corredor, John R. Wieczorek, Ricardo Ortiz Gallego & Leonardo Buitrago (2021).
- Using Google Refine and taxonomic databases (EOL, NCBI, uBio, WORMS) to clean messy data
  iPhylo blog post. Rod Page 2012.
- Reconciling author names using Open Refine and VIAF
  iPhylo blog post. Rod Page 2013.
- Validating scientific names with the GBIF Portal web service API
  Guest post was written by Gaurav Vaidya, Victoria Tersigni and Robert Guralnick 2013.
- iDigBio Cleaning data with OpenRefine
  iDigBio 2013.
- Have We Got the Names “Right”?
  Canadensys 2014.
- Cleaning data with OpenRefine
  Desmet and Brosens 2016 TDWG.
- EasyOpen Redlist
  Querying the IUCN Red List, using a species list, OpenRefine, and some pre-written code. Olly Griffin July 2019.

規劃／協作

- Agile methodology
- What is SCRUM
- SCRUM Framework
- Kanban methodology

- Scrum Guide
- GitHub

品質

- Principles of Data Quality
  Arthur Chapman 2005.
- Principles and Methods of Data Cleaning: Primary Species and Species-Occurrence Data
  Arthur Chapman 2005.
- Be careful with dates in Excel
  DataOne 2014.
- Character encoding for beginners
- MVZ Guide for Recording Localities in Field Notes

敏感物種

- Current Best Practices for Generalizing Sensitive Species Occurrence Data
  Arthur D. Chapman 2020.

分類學

- GBIF checklist datasets and data gaps
- GBIF Labs - Names Parser
- GBIF Labs - Species Matching
- Global Names Resolver
- Marine Name Matching Strategy for taxonomic quality control
- Nomenmatch

詞彙表

ALA

澳洲生物誌：為全球生物多樣性資訊機構（GBIF）的澳洲節點，開發了開源資料入口網站，目前大多被 GBIF 社群和合作夥伴用於建立國家級生物多樣性入口網站。

API

應用程式介面：一套明確定義的軟體元件間溝通方法。

BID

生物多樣性資訊發展計畫：由歐盟資助、GBIF
協調的專案，目標是提升非洲、加勒比海和太平洋地區的資料流通能力。

BIFA

亞洲生物多樣性基金。

CC Licences

創用 CC 授權條款：是由 CC 公眾授權組織（Creative Commons）
建立的一系列授權條款，透過提供免費的法律工具來促進創意和知識的分享與再利用。GBIF
分享的資料集可以使用其中三種授權方式：CC0、CC BY 和 CC BY-NC。

Controlled Vocabulary

「控制詞彙」是用於特定欄位的可能值的受限制詞彙集，可以將其視為特定欄位的查詢清單或下拉選單。例如，達爾文核心標準（Darwin Core）欄位的紀錄類型（basisOfRecord）應只包含以下值之一："PreservedSpecimen"、"FossilSpecimen"、"LivingSpecimen"、"HumanObservation"、"MachineObservation"。我們稱這個值列表為該欄位的控制詞彙。

DwC

「達爾文核心標準」是一個生物多樣性資料標準，由 TDWG 維護，在 GBIF 社群和合作夥伴中被廣泛使用。它是一組標準化的術語（或欄位名稱）及其定義，用於分享生物多樣性資訊。

DOI

數位物件識別碼：用於唯一識別物件的永久識別碼。DOI
主要用於識別學術、專業和政府資訊，如：期刊文章、研究報告、資料集和官方出版品。

DwC-A

達爾文核心文件包：一個壓縮（zip）檔案，包含分享給 GBIF 的特定資源所需的所有資訊。每個壓縮檔包含三種類型的檔案：

1. 實際資料（一個或多個文字檔）：occurrence.txt/event.txt/measurmentoffact.txt 等
2. 對應檔案：rtf.xml
3. 詮釋資料（EML）檔案：eml.xml。當你使用 IPT 發布時，它會建立一個 Darwin Core Archive，並與 GBIF 分享。此外，當你從 GBIF 網站下載資料時，也可以選擇 DwC-A 格式。

GUID

全球唯一識別碼

IPT

資料整合發布工具：這是一個免費開源的網路應用程式（軟體），用於發布生物多樣性資料。該軟體必須存放在具有 24/7 網際網路存取的伺服器上（可以是你的機構或其他地方）。它用於建立和處理可供任何人（包括 GBIF）分享和使用的 Darwin Core Archive 檔案。

Loan

借閱：在自然史收藏的語境中，這是指機構間標本借閱的程序。

LSID

生命科學識別碼：它們是生物物件的永久性全球唯一識別碼。

Data Publishing

「資料發布」在 GBIF 中有非常具體的定義。它指的是透過存取點（通常是網址 URL）以標準化形式使生物多樣性資料集可公開存取和發現。

Resource

資源：用於指稱已上傳到 IPT 實例的特定資料集及其詮釋資料的集合術語。

TDWG

分類資料庫工作組：現已更名為生物多樣性資訊標準（Biodiversity Information Standards）。

URN

統一資源名稱

UUID

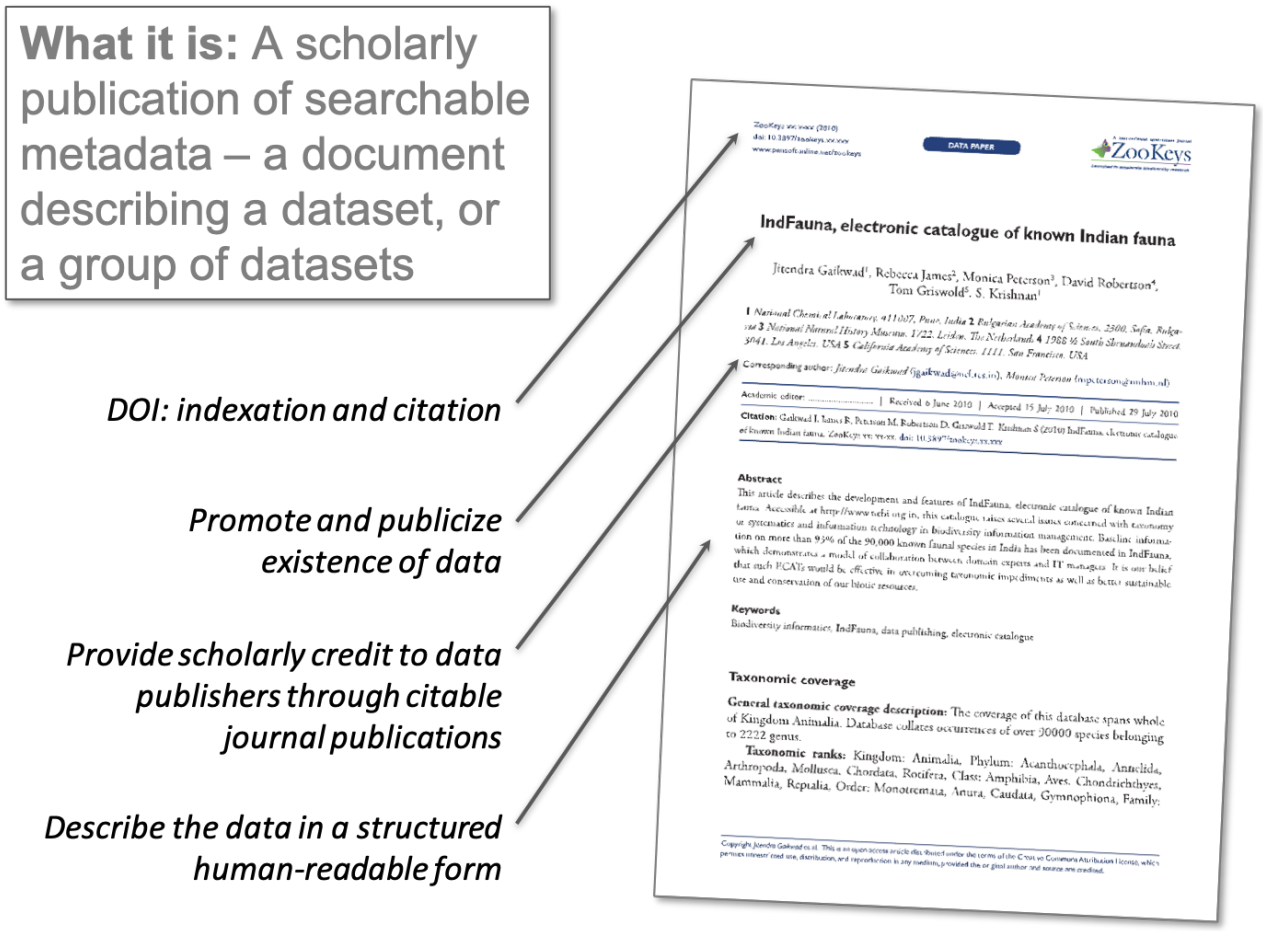
通用唯一識別碼

附錄：資料論文

資料論文是描述資料集的同儕審查文件，發表在同儕審查期刊上。準備、管理和描述資料需要付出努力。資料論文透過學術文章的形式對這種努力給予認可。我們在本課程中不會介紹如何建立資料論文，但作為選擇性活動，你可以觀看 Lizanne Roxburgh 主講的影片（51:51）。在這個影片中，你可以了解更多關於發布資料論文的資訊。如果你無法觀看嵌入的影片，你可以下載到本機端觀看。（MP4 - 99.2 MB）

▶ https://vimeo.com/265350948 *(Vimeo video)*

你可以在 GBIF.org 上閱讀更多關於資料論文的資訊。



點擊下方連結可以看到資料論文在 IPT、GBIF.org 和期刊上的呈現方式。這些都是交叉連結的。

- 期刊：https://doi.org/10.3897/BDJ.2.e4244
- GBIF: https://www.gbif.org/dataset/3c9e2297-bf20-4827-928e-7c7eefd9432c
- IPT: http://130.120.204.55:8080/ipt/resource.do?r=harvestmen_of_french_guiana

附錄：解答

本附錄包含所有複習測驗的答案和補充資訊。此外，本節還包含 USE CASE I 的建議解答。

基礎知識複習解答

針對給定的敘述，填入正確的術語（資料庫、資料庫語言、資料庫程式）

- 在統一介面中結合並呈現操作資料的功能和特性
  資料庫程式
- 在電腦上保存結構化且有組織的資料和／或資訊集合
  資料庫
- 人類與電腦溝通的方式
  資料庫語言

如果**你**開**啟**資料**檔**案時看到以下情況，**你**會懷疑是什**麼**問題？

□tre, ou ne pas □tre, c□est l□ la question.

- 使用了錯誤的編碼來開啟檔案

對於指定的軟體，請輸入軟體類型（資料**擷**取、資料管理、資料清理、資料發布）

- 整合式資料發布工具（IPT）
  資料發布
- 具體說明
  資料**擷**取 和 資料管理
- 愛自然（iNaturalist）
  資料**擷**取
- OpenRefine
  資料清理

對於提供的範例，請輸入正確的資料類型（二進位、布林**值**、浮點數、整數、長整數、文字、非結構化文字）

- 1236975
  長整數
- 01101111
  二進位
- 我們從鎮中心的郵局往西走了 5 英里。然後，我們沿著泥土路往北走了 2 英里到河邊。接著，我們沿著河繼續往西走了 5 英里。
  非結構化文字
- 1024
  整數
- 29.0
  浮點數
- Yes/No
  布林**值**
- 觀察到 6 隻兔子
  文字

哪些詞語描述了「欄位／欄位名稱」？

- 已分配
- 辨識中
- 獨特

哪些詞語描述了「欄位標籤」？

- 描述性
- 可讀性
- 使用者介面

對於**每**個陳述，輸入正確的結構（列、欄、表格）

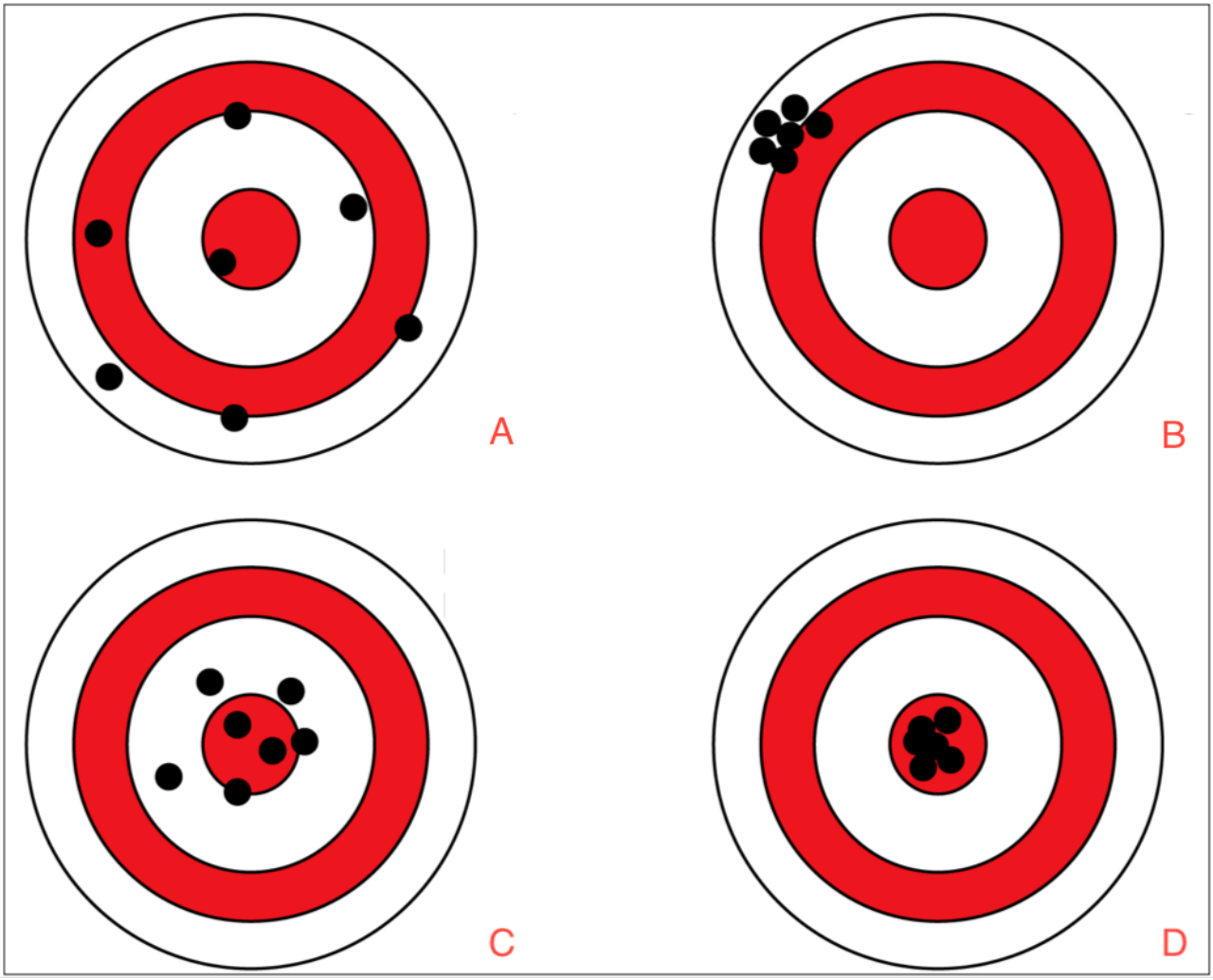
- 所有資料都指向單一概念。
  表格

- 一個屬性對每個記錄都有相同的欄位／資料類型。
  欄
- 記錄的屬性總是保持在一起。
  列

誰決定你的資料使用適切性？

- 用於研究或教育的資料使用者

For the given statements, input the matching image. (A, B, C, D)



- 高正確性，低精準度
  C
- 低正確性，高精準度
  B
- 高正確性，高精準度
  D
- 低正確性，低精準度
  A

辨識需要將資料集 **B** 合併到資料集 **A** 的資料關係（**0:1**、**1:0**、**1:1**、**1:∞**、**∞:1**、**∞:∞**）。並非所有關係都使用。

- 收集者欄位同時存在於資料集 A 和 B 中
  1:1

- 國家欄位只存在於資料集 B 中
  0:1
- 名稱欄位存在於資料集 A 中，但資料集 B 包含名字和姓氏欄位
  1:∞
- ID 欄位同時存在於資料集 A 和 B 中
  1:1
- 海拔存在於資料集 A 中，但不在資料集 B 中
  1:0
- 日期存在於資料集 A 中，但日、月、年在資料集 B 中是獨立欄位
  1:∞

詮釋資料很重要是因為（選擇正確的陳述）

- 它讓使用者能夠確定資料集是否適合他們使用。
- 它讓你知道在什麼法律條款下允許重複使用資料。

計畫檢查辦法

專案管理知識體係指南（**PMBoK**）五個流程群組的順序是什麼？

- 起始、規劃、執行、監控、結束

想知道更多：https://quizlet.com/306742513/1-introduction-pmbok-guide-6th-edition-flash-cards/

可交付成果的型態是什麼？

- 明確的 - 是
- 隱含的 - 是
- 估算的 - NO
- 直接的 - YES
- 直接的 - YES
- 猜想的 - NO

什麼是瓶頸？

- 使推進、進程延遲的阻礙 - YES 因某人事物缺失而造成的空白 - NO, 那是GAP
- 阻止某人做某件事、甚至使其不可能的問題與狀況 - NO , 這稱之為BARRIER

Which are examples of mobilization tasks?

- Affiliation - NO, This is a Resource Type
- Publishing - YES
- Imaging - YES
- Georeferencing - YES
- Increased Public Awareness - NO, This is an implied goal.

Data capture review solutions

What dataset type(s) would you choose for an ichthyology collection?

- occurrence
  Most of the time, specimens from collection databases are shared as occurrence data. Each occurrence (specimen or group of specimens) has its own unique identifier (sometimes derived from its catalogue number in the source collection) and the Darwin Core fields used to share them within GBIF describe each specimen: scientific name, the date it was collected on the field, who collected and/or identified it, where, etc. Each collection can have more than one specimen from a same species, as long as each specimen is identified by a unique ID.
- checklist
  It is also possible to create and share a taxonomical checklist derived from a collection database; in this case, it is recommended to share the checklist as a taxonomical dataset, with the occurrence (specimen) list associated with it by using the Occurrence core as an extension to the Taxon Core on the GBIF IPT.

What dataset type(s) would you choose for a list of invasive species?

- occurrence
  Some data publishers will share occurrence datasets coming from studies or programs tracking specimens from some specific invasive species; when the data focuses on individuals instead of the invasive species, in general, they can be shared as occurrence data.
- checklist
  Invasive species can be tracked and monitored at different scales (regional, national, thematic…); as this type of dataset focuses more on the species and their distribution across a given geographical scope, they are mainly shared as taxonomical datasets within GBIF (see GRIIS search results).

What dataset type(s) would you choose for the flora and fauna of an environmental impact study?

- occurrence
  Data are recorded by naturalists on the field and can be shared as simple occurrence datasets.
- sampling event
  They can also be shared as event datasets if standardized protocols (such as vegetation plots, transects, traps…) are used to collect the data.

What dataset type(s) would you choose for bird tracking data?

- occurrence
  These data are shared as occurrence datasets: ideally, each bird is identified with its organismID, and each occurrence (GPS ping) has its own occurrenceID, which is useful to track the different GPS locations of the same bird over the scope of the tracking programme or project. (See example)

What dataset type(s) would you choose for insect trap data?

- occurrence
  Although such data can be shared as simple occurrence datasets, it is best if they're shared as event datasets, where the location, identifier and contents of each trap can be better detailed.
- sampling event
  Insect traps (as well as other traps such as pitfall traps, malaise traps…) are typically used in

monitoring programmes to check the presence (or absence) of some species and/or assess their specific abundance. Using the "eventID" field to identify each trap allows the users to get all of the specimens collected within each trap. The same logic applies to other field protocols such as transects, plots, remote cameras, etc.: by using the Event Core instead of the Occurrence core, you'll be able to share much more information about the context of the data collection, and allow users to better understand (and even replicate) your work.

What dataset type(s) would you choose for national park management data?

- occurrence
  record individuals of species
- checklist
  It is important to know how many species are present in the park/reserve perimeter and their conservation status.
- sampling event
  check and track the populations

What dataset type(s) would you choose for a citizen science bioblitz?

- occurrence
  Bioblitz datasets are mainly shared as occurrence datasets.
- sampling event
  Depending on the citizen science programme, specific sampling protocols might be used by the volunteers, in which case, the data can be shared as an event dataset.

What dataset type(s) would you choose for a regional species list?

- checklist
  Geographical or thematic species lists are often used to share information about the species present in a given area; most of the time, these lists also mention the distribution of each species as well as their conservation status in this area. Regional species lists can give a useful insight into a region's biodiversity and habitats, and need to be shared as taxonomical datasets, with or without associated occurrences.

Data management review solutions

Why is it best to clean your data?

- to make them as fit for use as possible
- to achieve your data quality goals

You should always aim to manage and publish data with the highest possible quality. This will improve your day-to-day work (it is easier to work with organized and clean data), as well as the work of potential re-users of your data, who need to understand them and trust their source before using them.

How should you organize your data cleaning workflow?

- ask your colleagues for expertise
- work at an institutional level to harmonize data quality workflows

Nobody is expected to know everything about biodiversity data; you should seek help and advice from your colleagues or other knowledgeable people, and ensure that you're applying the good practices

recommended by your institution as you clean your data.

Which is best:

- prevent errors from occurring
- correct errors as soon as you find them in your database or spreadsheet

The best way to avoid spreading errors in your data is to prevent them from occurring at the start of the data collecting/recording process.

Of course, mistakes are unavoidable so you should also clean them as soon as you find them, and document the cleaning process.

If you don't have the time or resources to properly clean your data, it is best to wait before you can do so instead of publishing erroneous data that might confuse people.

Whose responsibility is data quality?

- Everyone involved in the management of data

Every person involved in your data management workflow is at least partly responsible for their quality, from the field technicians to the database manager(s).

People who might later use your data can inform you of any remaining error in your data, and should use them responsibly for their own research, but the initial data quality is not their responsibility.

GBIF can perform automatic checks on your data (e.g. detection of missing values, geographic outliers, unknown scientific names) but should not be held responsible for errors that occurred earlier in the data management process.

下列**哪**些工具可以用來清理**你**的資料？

- Excel & other spreadsheets management tools
- OpenRefine
- Your database software
- Online tools such as Scientific Names Resolver or Google Maps

All kinds of tools can be used to clean your data, but you should identify which ones will answer your needs in terms of taxonomic resolving, georeferencing, deleting duplicates, and so on. You can find helpful tools listed in the data management section.

Data publishing review solutions

What does data publishing mean in the context of GBIF?

- Making your biodiversity dataset(s) publicly accessible and discoverable in a standardized format

Data publishing within GBIF means making your biodiversity dataset(s) publicly accessible in a standardized format (most of the time, Darwin Core), so that it can be discovered and reused by other people.

What is an IPT?

- a tool that helps you publish your data to GBIF

- a tool that helps you produce a Data paper

The IPT (Integrated Publishing Toolkit) is a Java-coded software that allows you to upload and publish data to GBIF. It is not to be used as a data management or data cleaning tool.

The IPT can also help you with the process of writing and submitting a data paper, thanks to the EML file it generates automatically when you fill in the metadata for your data resource.

Which Creative Commons licences and waivers are recommended by GBIF for data publication?

- CC0, CC-BY and CC-BY-NC

The Creative Commons licences and waivers recommended to publish your dataset(s) to GBIF are CC0, CC-BY and CC-BY-NC. They are widely recognized licenses and/or waivers that align with international open-data requirements for data sharing and re-use.

Please note that you should only choose CC0 or CC-BY waiver/license for your BID-related dataset(s).

What are the three Cores from which you can choose for an IPT resource?

- Occurrence Core, Taxon Core, Event Core

You can choose one of the three following Cores for each of your IPT resources: Occurrence, Taxon or Event Core.

The Darwin Core standard also allows you to link extensions to your chosen Core, such as SimpleMultimedia or MeasurementOrFact.

The metadata are filled in a separate section of the IPT and are shared using the EML standard, not the Darwin Core (which is used for data only).

How many Extensions files can a dataset have?

- as many as needed

Once you have chosen a Core for your IPT resource, you can add Darwin core extensions to it. You can add only one or several extensions, depending on the type of Core you chose, and which extensions are compatible with it.

Extensions are not mandatory (you can publish a dataset without any extension) but can be useful if you want to share additional information that you could not map with your chosen Core.

Use Case I suggested solution

suggested solution (PDF 144 KB)

Acknowledgements

Course design and instruction

The success of this course depends heavily on the support provided to participants from GBIF's network of capacity enhancement mentors. Visit the GBIF page on capacity enhancement mentoring to read more about these individuals and their contributions.

The following individuals are recognized for their significant contributions to the course design, materials and instruction:

- Nestor Beltran*
- David Bloom
- Katia Cezón*
- Dag Endresen
- Alberto González-Talaván*
- Sharon Grant*
- Marie-Elise Lecoq
- Sophie Pamerlon*
- Nicolas Noé*
- Mélianie Raymond*
- Laura Anne Russell*
- John Wieczorek
- Paula Zermoglio

*Originators of the curriculum

Special acknowledgement to Arthur Chapman for the reuse of his materials on Data Quality.

翻譯者

法文

- Maxime Coupremanne
- Jaures Gbètoho
- Marie Grosjean
- Patricia Mergen
- Sophie Pamerlon
- Andry Jean Marc Rakotomanjaka
- Y. Sabastian Wirsiy

葡萄牙文

- Rui Figueira
- Clara Baringo Fonseca
- Keila Elizabeth Macfadem Juarez
- Tainan Messina

西班牙文

- Leonardo Buitrago
- Victor Chocho

- Camila Plata
- Anabela Plos
- William Ulate
- Paula Zermoglio

Resources

- The online tabletop platform is provided by PlayingCards.io. Much appreciation and thanks to Jwalant Patel and Eric Ma for finding and helping to create the online playing tables and to Kate Webbink for artistic expertise.
- 此練習中的圖示(Icons)來自 Freepik from www.flaticon.com
- OpenRefine
- Integrated Publishing Toolkit

Resource support

- Belgium Biodiversity Platform
- GBIF France
- GBIF Norway
- GBIF Spain
- SiB Colombia
- The Field Museum
- VertNet

Colophon

Suggested citation

GBIF Secretariat (2021) GBIF Biodiversity Data Mobilization Course. 12th edition. GBIF Secretariat: Copenhagen. https://doi.org/10.35035/ce-c6cr-6w42. [Date of course.]

Contributors

The *GBIF Biodiversity Data Mobilization Course* was originally developed as part of Biodiversity Information Development, a programme funded by the European Union. The original curriculum was created by Nestor Beltran, Sharon Grant, Nicolas Nóe, Sophie Pamerlon, Alberto González-Talaván, Mélianie Raymond, Laura Anne Russell and Katia Cezón, with additional contributions by GBIF trainers, mentors and students.

Licence

Persistent URI

https://doi.org/10.35035/ce-c6cr-6w42

Document control

12th edition, May 2021