

發布物種名錄最佳指南

Publishing Species Checklists

Best Practices

Version 1.11



2011 年 4 月

原文建議引用格式如下：

GBIF (2010). Best Practices in Publishing Species Checklists, (contributed by Remsen D., Döring M., Robertson, T.), Copenhagen: Global Biodiversity Information Facility, 20 pp, accessible online at http://links.gbif.org/checklist_best_practices

中文建議引用格式如下：

鄧惠瑜、林思吟（譯）（2011）。發布物種名錄最佳指南。台北：中央研究院生物多樣性研究中心。（GBIF (2010). Best Practices in Publishing Species Checklists, (contributed by Remsen D., Döring M., Robertson, T.), Copenhagen: Global Biodiversity Information Facility, 20 pp, accessible online at http://links.gbif.org/checklist_best_practices）

本著作透過達爾文核心資料集(一種國際資料交換格式)提出共享物種名錄的建議。

國際標準圖書編號：87-92020-26-7

永久網址：http://links.gbif.org/checklist_best_practices

語言：中文

版權 © 全球生物多樣性資訊機構，2010

授權條款：



本著作係採用創用CC 3.0 聲明頁面授權條款授權。

版本更新：

版本	描述	發表日期	作者
1.0	發行版本	2011年4月1日	David Remsen
1.1			David Remsen
1.11	些微編輯變更	2011年4月28日	David Remsen

封面藝術：Gregory Basco

黑頸長腳鷸 *Himantopus mexicanus*

聯絡GBIF

David Remsen (dremsen@gbif.org) 負責本著作之文章、格式及其原始檔案。

Markus Döring 負責附加的技術細節。

本著作也屬於「GBIF 資料發布手冊1.0版，ISBN 87-92020-31-3」的一部分，網址：

http://links.gbif.org/data_publishing_manual

發布物種名錄最佳指南

本文翻譯目的是在台灣或其他華文地區推展生物多樣性原始資料的整合，如對於中文內容有任何意見，懇請與出版單位聯繫。

本文件出版由國科會支助之「台灣生物多樣性資訊機構」及「數位典藏與數位學習國家型科技計畫」提供。



數位典藏與數位學習 國家型科技計畫
Taiwan e-Learning and Digital Archives Program

發行人 李文雄
企劃 邵廣昭
總監 賴昆祺
翻譯 鄧惠瑜 林思吟
校對 邵廣昭
編輯 李香瑩

出版單位 中央研究院 生物多樣性研究中心

地址：115 臺北市南港區研究院路二段 128 號

電話：02-27872221

全球資訊網：<http://biodiv.tw>

台灣生物多樣性資訊機構 <http://taibif.org.tw>

出版日期 中華民國 100 年 4 月 1 日 初版

GBIF 簡介

全球生物多樣性資訊機構（以下簡稱GBIF）是為了因應21世紀的重要挑戰之一－治理地球的生物多樣性知識－而成立的全球性大科學倡議機構。GBIF的願景是「一個在科學、社會及永續未來皆能自由和普及地取用生物多樣性資訊的世界」。GBIF的使命：成為全球最重要的生物多樣性資訊資源，促成環境和人類福祉的明智解決方案。為了達成使命，GBIF鼓勵全世界各式各樣的資料發佈者，透過GBIF網絡一起來發現及發布資料。

¹ GBIF (2011). GBIF 策略計畫 2012-16：掌握未來。哥本哈根：全球生物多樣性資訊機構。第7頁。 ISBN: 87-92020-18-6。網址 http://links.gbif.org/sp2012_16.pdf

目錄

GBIF 簡介	iv
Introduction 前言	1
Scope 範圍	1
Darwin Core Archive Format 達爾文核心集檔案格式	3
<i>Checklist Metadata</i> 物種名錄後設資料	3
<i>Checklist Data</i> 物種名錄資料	3
<i>Data file formatting recommendations</i> 資料檔案格式化的建議	4
Sharing Scientific Names 共享物種學名	5
A. Concatenated in the scientificName field 學名欄位串聯	5
B. Separate Name and Authorship parts 物種學名與命名者姓名分開	5
C. Separated into name parts 物種學名拆解成多個部分	6
Infrageneric Markers 屬以下的分類範圍標籤	6
Publishing Classifications 發布分類資料	7
<i>Normalised Classification (Parent/Child)</i> 標準化分類 (親/子)	7
Advantages 優點	8
Disadvantages 缺點	8
<i>Denormalised Classifications</i> 非標準化分類	8
Advantages 優點	9
Disadvantages 缺點	9
Other classification-related recommendations 其他相關分類建議	9
Classification Formats <u>not</u> recommended for Publishing 不建議的發布分類格式	10
Publishing Synonymy 發布同種異名	11
<i>Nomenclatural Synonymy</i> 命名上的同物異名	13
<i>Pro-parte Synonymy</i> 部分同種異名	13
Citation and Attribution 引用和歸屬	13
Metadata Citation and Attribution 後設資料引用和歸屬	14
Data-level Citation and Attribution 資料層級的引用和歸屬	15
Authenticated File Access via httpd 透過 httpd 身分驗證存取檔案	15
Use Case #1 – Checklist composed of multiple contribution datasets (e.g., Catalogue of Life, PESI, WoRMS) 使用案例 #1 - 由多個資料集組成的名錄 (例如: Catalogue of Life, PESI, WoRMS)	15
Use Case #2 – Checklist derived from one or more authority sources 使用案例 #2 – 從一個或多個授權來源獲得的物種名錄	17

Sharing Vernacular Names 共享俗名	18
Sharing Species Descriptions 共享物種描述	18
Multi-line descriptions 多段描述	19
Sharing Species Distributions 共享物種分布	19
Sharing References 共享參考文獻	19
Sharing Type information 共享模式標本資訊	20
Sharing Links and Identifiers 共享連結及識別碼	20
Creating a dynamic link to species page 創造物種頁面的動態連結	20

Introduction 前言

本指南詳述如何運用達爾文核心集檔案(DwC-A)格式作為工具，在一個標準方式下共享分類學名錄資訊。本文著重在達爾文核心集檔案格式的特定組件，和一些核心分類資料層級（core taxonomic data class）的支持延伸，並建議如何最有效地利用這些組件，以將共享資料的價值最大化。本文不專門提供達爾文核心集檔案格式的詳細介紹，也不是描述全球名稱體系概況 GNA Profile 完整的專門用語集和延伸附錄的詳細參考指南。另有兩份文件包含上述的主題：

1. Darwin Core Archive How to Guide²（達爾文核心集檔案指南）
2. GNA Profile Reference Guide³（全球名稱體系概況參考指南）

本文所描述的達爾文核心集檔案格式以及特定介紹，對共享的分類資料而言，代表一種受到國際公認和批准的資料交換格式。所有交換資料標準都必須同時在技術範圍和能力、以及在社會認可和接受度之間取得平衡。簡單的解決方案為了使用的方便，而犧牲了資料的涵蓋範圍及複雜度；高度複雜的格式為了各種類型資料提供較完整的解決方案，但是缺少了簡單性，還需要軟體及專門技術的支援。達爾文核心集檔案格式是兩個極端之間的折衷辦法。它側重在分類名錄中的關鍵要素，使豐富的資料類型集能夠連接到這個核心結構。熟悉基本結構化文字檔案（basic structured text files）的生物學者和資料管理者，能夠很容易地理解及使用已收錄歸檔的資料。透過提供一個比較容易生產、消費，並且支援組成多種分類資料資源庫關鍵要素的國際標準，GBIF 希望提供標準化的途徑給物種名錄的創造者（creators）及管理者，以分享他們的資料，並促進後續引用及認可他們貢獻的通用辦法。格式的標準化，也能增加關連性及利用。

² http://links.gbif.org/gbif_dwca_how_to_guide

³ http://links.gbif.org/gbif_profile_reference_guide

Scope 範圍

物種名錄“species checklist”與分類學的目錄“catalogue”可能涵蓋了分類學資源上的重疊範圍。所有學名集合的產出，皆明確或隱含地提到分類單元。名錄中的名字組，常受限於分類群、地理位置、或是主題(如入侵種)，甚或是包含以上三種的組合約束。為了增加更多細節的詮釋，本文將名錄區分為以下類型來討論⁴

1. Name lists 名錄 – 簡單的物種名字清單，沒有明確的指出分類學地位，但是通常是指已被認可的分類群名稱。這類的名錄通常來自於某個地區或主題內容裡的一份分類群清單。
2. Nomenclatural list (Nomenclators) 命名名錄（命名者、詞彙手冊） – 包含名字的分類單元，是依照相關的命名編碼(Codes of Nomenclature)，來註冊已經發布使用的學名。大多數命名行為（acts）是指新學名的原始描述（original descriptions），亦可能包含校正（emendation）、選定模式（lectotypifications），以及其在他法規所規範的行為。以分類的概念來看，同種異名並不包含在這個名錄中，而是(為了植物學家們)以新建的關聯連結至基礎名(basionym)，提供命名上的同物異名。
3. Taxonomic checklist 分類學名錄 – 此類名錄藉由增加明確的分類地位以及包含同種異名，從「命名名錄」擴充為「分類學名錄」。簡單的分類學名錄對於同種異名並不提供明確定義細節。分類單元通常單純以物種間相似性比較，呈現有組織的分類（Taxa are organized into classifications）。分類學目錄“taxonomic catalogue”以及其餘類型也可用於這樣的例子，特別是包含合法發表的出版物及分類地位等細節的資源。
4. Annotated Checklists 已註記名錄 – 這個類型是由「分類學名錄」擴充成「已註記名錄」，透過加入其他資料形式（註記）到核心文件、同種異名名錄，例如俗名、受威脅等級、分布及基本描述資訊等。當註記欄位提供足夠的資料來有效的定義、或界定一個分類群時，例如詳細的特徵描述和插圖、分子資料、標本等等，這個「已註記名錄」才可以歸類為以下定義的兩個類型其中之一。

⁴ 這些類型及描述是直接引用自 “Hyam . R., Standardisation of Data Exchange in the Pan-European Species-directories Infrastructure (PESI) Deliverable D 4.1”

5. Flora or Faunal list 植物或動物名錄 – 這類名錄為提供某一特定地區詳細物種說明的典型書籍。細節包含了「已註記名錄」中多種資料類型，例如詳細文字描述、插圖、標本參考資料，以及其他在該地區但不一定是全球明確定義（界定）的分類群詳細資料。
6. Monographs 專論 – 專論是詳盡的物種描述，通常是針對某特定分類群的全球性尺度詳述專書。專論包含詳細的命名、同種異名和分類單元定義，包含文字描述、插圖、引證標本資料和參考文獻的書目。

達爾文核心集檔案格式與全球名稱體系概況(GNA profile)支援以上所有名錄類型裡關鍵資料元素的交換，具體的資料覆蓋度取決於提供資料的每一個原始資源。在本文中，以廣義的物種名錄“checklist”代表上述所有或任一個資源類型。當特定的資源被引用時，便會使用相對應的特定類型。

Darwin Core Archive Format 達爾文核心集檔案格式

達爾文核心集檔案 (DwC-A) 是一個資訊學資料標準，它利用達爾文核心集語法，為物種名錄資料產生一個單獨且自成體系的資料集。在檔案庫裡，檔案的集合形成一個自成體系的資料集，這個資料集能以單一個壓縮檔 (Zip 或 GZIP) 的形式提供給使用者。資料集是由描述式後設資料文件和一組一或多個資料檔案所組成。

Checklist Metadata 物種名錄後設資料

為了經由 GBIF 網路發布物種名錄資料，必須先將資料集中的出處和範圍文件化。資料集文件是一種「後設資料資源」 (resource metadata)，以便使用者評估資料集的適用程度。資料集文件詳述名錄的範圍、功能、方法學、編輯時所需的資源、以及相關的創造、管理之個人與機構。後設資料是以達爾文核心集檔案共享的 XML 文件。GBIF 提供以生態學後設資料語言 (Ecological Metadata Language) 為基礎的物種名錄後設資料概況 (metadata profile)。使用指南詳述了利用本格式描述物種名錄的所有選項，請參見

http://links.gbif.org/gbif_metadata_profile_how_to_en_v1

Checklist Data 物種名錄資料

達爾文核心集檔案格式提供了發布物種名錄的框架。達爾文核心集檔案是由一個或多個文字檔案串連所組成，其文字檔案是用逗點或 Tab 分隔的標準格式。文件檔案是以星狀法邏輯排列，中間是核心檔案，包含基本名錄要素（物種清單、分類、同物異名），周圍則是數個擴充檔，以描述相關的資料類型（例如俗名）。中心檔及擴充檔之間的連結是利用分類單元識別碼（taxonID）資料元素建立的。依照這個方式，每一筆核心分類記錄可以並存多筆擴充記錄。這樣的星狀圖“star-schema”提供一個簡單的相關資料模型，以支援許多物種名錄共同的註記型態。

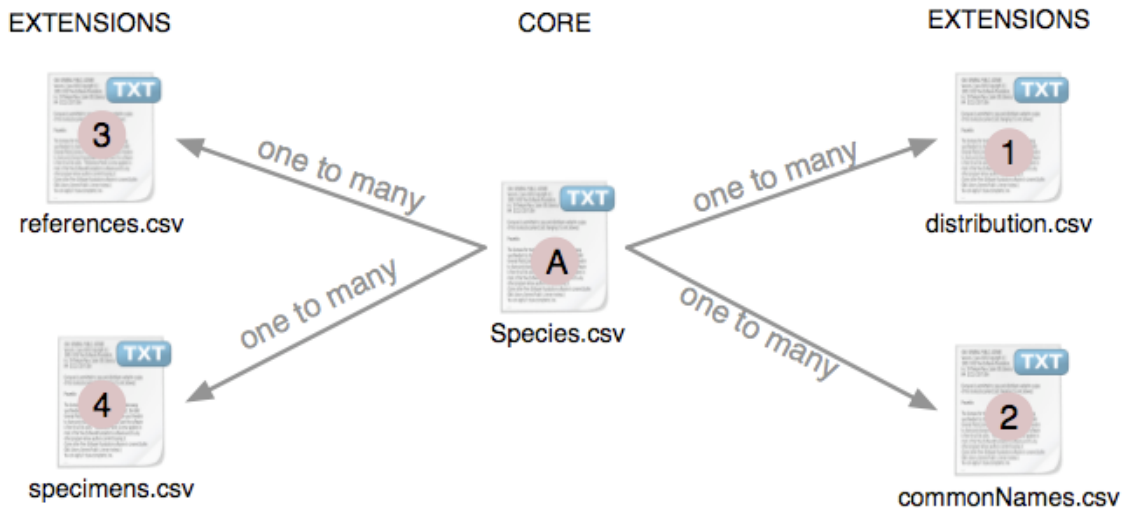


Figure 1 - Darwin Core Archive data files in 'star schema'

EXTENSIONS 擴充檔案，CORE 核心檔案，one to many 一對多。

圖 1-以「星狀圖」呈現達爾文核心集檔案資料

Data file formatting recommendations 資料檔案格式化的建議

為了有助於理解，本指南中，以欄位(field)代表分類學發布概況裡的達爾文核心集語言組，以標示使用者資料的狀態。舉例來說，用「dwc:scientificName」欄位來代表「達爾文核心集：學名(scientificName)」。

- 建議使用 Tab 或逗點分隔數值 (Comma-Separated-Values) 的方式取代一般用來定義欄位 (custom field) 的分隔符號 (delimiters) 及引用符號 (quotes)。
- 小心謹慎，並維持引用的一致性。
- 以 UTF-8 編碼文字檔案。
- 確認取代數據域中所有的分行符號 (line breaks)，例如 `\r \n` 或有單空格的 `\r\n`、或是想刻意保留的話，可使用 2 個字符 (characters) 如 “`$$`” 來取代 “`\r`”，以避免分行。另一個選項是用 html 語言的 `
` 標籤來取代分行符號。

- 將 NULL 設定為空串(empty strings)的編碼,例如 2 個分隔符號中間沒有字符(characters)時,或是用 \N 或是 \NULL,而不再加其他的文字序列。

本指南不加詳述達爾文核心集檔案格式,請連結此網址⁵獲得達爾文核心集檔案格式指南一文。

Sharing Scientific Names 共享物種學名

達爾文核心支援多種共享物種學名的方法,包含以下選項:

A. Concatenated in the scientificName field 學名欄位串聯

scientificName
Gerardia paupercula var. borealis (Pennell) Deam

在 dwc:scientificName 欄位中,儲存著包含命名者姓名的完整學名。這個欄位應該永遠填充著數據,即使學名被拆解成(如下述 C 選項)的形式也應該存在。資料庫中沒有將命名者姓名與學名清楚分開時,應該要將全部的名字字串填入此欄位中。使用雜交命名規則(hybrid formulas)命名的名字、狹義上的名字(sensu stricto names)、自動名(autonyms)、非平凡二名法的名字(non-trivial binomials)都可填入此欄位。這個欄位通常與 dwc:taxonRank 欄位一起使用,以儲存包含高階分類群的完整分類學名錄。

B. Separate Name and Authorship parts 物種學名與命名者姓名分開

scientificName	scientificNameAuthorship
Gerardia paupercula var. borealis	(Pennell) Deam

有些資料庫中將學名與命名者姓名分隔開,則可使用 dwc:scientificName 欄位與 dwc:scientificNameAuthorship 欄位。

⁵ http://links.gbif.org/gbif_dwca如何指南en_v1

C. Separated into name parts 物種學名拆解成多個部分

Genus	specificEpithet	taxonRank	infraspecificEpithet	scientificNameAuthorship
Gerardia	Paupercula	var.	borealis	(Pennell) Deam

達爾文核心集提供一系列的語法，使物種學名能拆解成各個單元。有些資料庫是依此解析單元的方式儲存物種名錄，在這種情況下，就可以考慮以這種形式共享資料。然而，如果依照 C 選項將學名拆解，仍強烈建議額外增加一個完整的物種學名欄位，並以 `dwc:scientificName` 字串來分享（如上述 A 選項所示）。注意在上列表格中，雖然未顯示達爾文核心用語 `dwc:subgenus`，但是它代表另一個組成部分學名的可能。

Infrageneric Markers 屬以下的分類範圍標籤

如果可能的話，請將屬以下的分類範圍標籤納入學名的一部分，以避免混淆屬以下分類階層與基礎名作者（basionym author）。以 “*Ageratina subgen. Apoda* R. M. King & H. Rob” 與 “*Ageratina (Apoda)* R. M. King & H. Rob” 為例，前者是較好的寫法，後者的 *Apoda* 可能是指亞屬（subgenus），也可以是指基礎名作者（basionym author）。

Publishing Classifications 發布分類資料

達爾文核心提供發布分類 (classifications) 或分類階層 (taxonomic hierarchies) 的兩個基本選項：標準化分類 (normalized) 與非標準化分類 (denormalised)。大部分資料庫都是使用這兩個選項進行分類管理。

Normalised Classification (Parent/Child) 標準化分類 (親/子)

共享分類資訊時，較推薦的方式是使用標準化分類格式，在資料庫中也稱之為親子關係 “parent-child relationship” 或鄰接表 “adjacency list”。在一個標準化的分類階層中，每一個分類階層單獨一行。這種分類層次包含了種名和更高階的分類階層。每一行至少都含有下列組成元素。

- *dwc:taxonID* 指當前分類階層的識別碼。
- *dwc:scientificName* 指當前的分類階層的學名。例如：虎 “*Panthera tigris*”
- *dwc:taxonRank* 指分類階層的位階。例如：種 “*species*”
- 分類階元的直接親級階元 (parent taxon) 的識別碼是放在 *dwc:parentNameUsageID* 欄位中。下面的例子裡，記錄 7-- 虎 “*Panthera tigris* (Linnaeus)” 的親級階元是記錄 6 的豹屬 “*Panthera*”。

以下是一物種的分類範例，虎 “*Panthera tigris*”。注意最頂端的分類階層是沒有親級階層的，因此它的親級階層識別碼應該是空白或是 “0”。注意在這個例子中 *dwc:scientificName* 是提供儲存名稱的一個共同欄位，但是一個完整學名填寫欄位在本文前面的章節 [Sharing Scientific Names 共享物種學名] 中有詳述。

taxonID	taxonRank	scientificName	parentNameUsageID
1	Kingdom 界	Animalia 動物界	0
2	Phylum 門	Chordata 脊索動物門	1
3	Class 綱	Mammalia 哺乳動物綱	2
4	Order 目	Carnivora 食肉目	3
5	Family 科	Felidae 貓科	4
6	Genus 屬	Panthera 豹屬	5
7	Species 種	Panthera tigris (Linnaeus) 虎	6

Advantages 優點

- 效率 – 標準化的分類將層級中的各分類階層儲存成單獨的參考欄位。
- 參照完整性 – 每個階層組成都有一個獨特的辨識碼，可以明確地顯示出直接的親級階元。這樣很容易檢驗分類層次的完整性和正確性。
- 擴展性 – 所有分類階層因為有獨特的辨識碼可以分辨彼此。可以透過與種的記錄相同的延伸方式，使得較高的階層有更詳實的紀錄。

Disadvantages 缺點

- 便利性 – 使用原始表格時，標準化的分類方法無法對分類階級提供一個直觀的視角。許多生物學家慣用效率較低的非標準化格式，視覺上看起來較直觀。要把非標準化的分類手動轉換成標準化的分類是相當困難的。

註： *dwc:parentNameUsageID* 必須指向一個已存在資料庫中現有的記錄。指向一個不存在的更高階層記錄的識別碼是無效的。

Denormalised Classifications 非標準化分類

對任何一位利用試算表 (spreadsheet) 來管理物種資訊的人來說，這種格式是非常熟悉的。在非標準化的分類格式裡，數據表的每一列都指向最終的分類階層，例如一個種名，同時每個親級的分類階層都在每個直欄位裡清楚呈現。

以達爾文核心集檔案共享分類學資料時，並不推薦使用這種方法。但由於非標準化分類格式在許多物種名錄中廣泛使用，因此 GBIF 也支持此格式。如果資料要以此方法共享，強烈建議：

1. 每一個較高階層欄位都要完全填滿。避免出現如以下範例表格中植物界部分欄位空白。
2. 確保分類列表的完整性。例如，必須確認兩個同屬的物種屬於同科。若同種異名被放在另外一列，要確認其對應到目前被接受之正確學名。

taxonID	kingdom 界	phylum 門	class 綱	order 目	family 科	scientificName*
1001	Animalia 動物界	Chordata 脊索動物門	Mamalia 哺乳動物綱	Carnivora 食肉目	Felidae 貓科	Panthera tigris 虎
1002	Animalia 動物界	Chordata 脊索動物門	Mamalia 哺乳動物綱	Carnivora 食肉目	Felidae 貓科	Panthera leo 獅
1003	Animalia 動物界	Arthropoda 節肢動物門	Insecta 昆蟲綱	Hymenoptera 膜翅目	Apidae 蜜蜂科	Apis mellifera 蜜蜂
1004	Plantae 植物界	--	--	Poales 禾本目	--	Poa annua 早熟禾

(譯註：原始表格第 4 行第 6 欄的 Poales 改放在第 4 行第 5 欄中。早熟禾學名 *Poa annua*，我推測 *Poa annularis* 可能是同種異名 *Poa annua* L. var. 的誤植。)

Advantages 優點

- 易讀性 – 這種格式主要的優點是它易於閱讀，僅靠閱讀欄位就可以評估分類的階層。
- 方便性 – 試算表的應用與其他相關的資料庫有助於方便地使用此架構來儲存分類階級資料。

Disadvantages 缺點

- 更可能喪失參照的一致性 – 在這種格式下，高階層的分類單元會不斷重複，這樣會增加兩個完全相同的分類階層拼字不一致的機會。這種格式也可能發生其他類似的風險，例如，兩個相同階層（例如貓科“Felidae”）有可能被分配到不同的親級階層裡，導致分類層級一致性的衝突。
- 缺乏較高階元的詳細資料 – 這種格式將較高階元視為一個物種的附屬，而非獨立的分類階層記錄。因此，無論是在核心檔案中或是在其他擴充延伸檔案，都無法將較高階層的記錄分享出去。

Other classification-related recommendations 其他相關分類建議

- 即使是基本的物種名錄，也應該將界（Kingdom）和對所有記錄的命名規約參考文獻包括進來。
- 在非標準化的分類裡，最起碼應該要包含界、門、科等分類資料。
- 如果整個資料庫都相同，考慮使用語法與數值的靜態對應（a static mapping of the term and value）。對於對應全球數值的詳細說明，請參考達爾文核心集檔案指南，網址：
http://links.gbif.org/gbif_dwca_how_to_guide_en_v1

Classification Formats not recommended for Publishing 不建議的發布分類格式

以下例子闡釋的幾個資料配置方式雖然符合概況，但是並不被推薦或不被 GBIF 所支持（也就是說 GBIF 的解碼器無法正確的處理這些例子）。

A. 這個例子列出從屬的分類階層，而最後一欄則是分類階層值。

taxonID	kingdom	phylum	Class	order	family	scientificName*
997	Animalia					
998	Animalia	Chordata				
999	Animalia	Chordata	Mammalia			
1000	Animalia	Chordata	Mammalia	Carnivora		
1001	Animalia	Chordata	Mammalia	Carnivora	Felidae	
1002	Animalia	Chordata	Mammalia	Carnivora	Felidae	Panthera tigris
1003	Animalia	Chordata	Mammalia	Carnivora	Felidae	Panthera tigris

B. 這個例子與上面 A 例相似，但是只記錄一次高階分類階元，試圖來降低出錯的可能，以維持資料的一致性。

taxonID	kingdom	phylum	class	order	family	scientificName*
997	Animalia					
998		Chordata				
999			Mammalia			
1000				Carnivora		
1001					Felidae	
1002						Panthera tigris
1003						Panthera leo

請避免用以上這些配置形式發布資料。

Publishing Synonymy 發布同種異名

達爾文核心集檔案支援同物異名的物種名錄發布。同物異名在核心資料檔案中是以獨立的記錄形式發布。同種異名利用 *dwc:acceptedNameUsageID* 欄位與已被接受的分類階層記錄互相參照。這個欄位包含 *dwc:taxonID*，代表被接受的分類階層記錄。在下述的簡化例子中，第一筆記錄代表已被接受的分類階層名字，記錄 2 和 3 則是同種異名。

taxonID	scientificName	acceptedNameID	taxonomicStatus	nomenclaturalStatus
1	Coeligena helianthea (Lesson 1838) 藍喉星額蜂鳥	1	accepted 合法名	
2	Ornismya helianthea Lesson 1838	1	Homotypic synonym 同型異名	
3	Helianthea helianthea (Lesson 1838) J. Gould 1848	1	Homotypic synonym 同型異名	
4	Helianthea typical Bonaparte 1850	1	Heterotypic synonym 異型異名	nomen dubium 疑問名
5	Helienthea porphyrogaster Mulsant 1876	1	Heterotypic synonym 異型異名	nomen dubium 疑問名
6	Coeligena helianthea tamai Berlioz & Phelps 1953	1	Heterotypic synonym 異型異名	nomen dubium 疑問名

同物異名記錄建議包含獨立的 *dwc:taxonID* 欄位，或完全拿掉 *dwc:taxonID*。已被認定的物種階層記錄不能使用重覆的 *dwc:taxonID*。最簡單的同物異名表示方法，就是如上面的範例一樣，將同種異名分別獨立的條列，同時利用 *dwc:acceptedNameUsageID* 欄位「指向」正式認定的分類階層命名。這種簡單的同物異名，能支援同物異名資料的基本名錄發表，其提供的細節又與核心分類階元層級元素 (core taxon class elements) 相符。*dwc:taxonomicStatus* 欄位確認了每一筆記錄的位階。對於該欄位的建議詞彙，可參考此網址⁶。額外的同物異名命名細節也可透過使用 *dwc:nomenclaturalStatus* 欄位和支援詞彙⁷ 來記錄。

⁶ http://rs.gbif.org/vocabulary/gbif/taxonomic_status.xml

每一筆同物異名的記錄，各自包含一獨立的 *dwc:taxonID* 欄位，以便提供更多同物異名的細節，並可藉由現有的擴充檔來支持物種名錄註記的分享。同時支持將一個或多個書目記錄、標本記錄和其他由全球名稱體系概況(GNA profile)提供的資料類型，連結至核心資料檔案中的單筆同物異名記錄。如果同物異名的記錄無法使用 *dwc:taxonID*，那麼依賴 *dwc:taxonID* 的擴充檔也不能連結用核心檔案中分類階層記錄。底下有一個簡單的案例來說明兩個文件(如表所示)的用法，此用法是藉由使用參考資料擴展 (Reference extension) 來提供一個同物異名的書目。共享的 *dwc:taxonID* 欄位在例子中以黃底標註。

Taxon.txt 資料文件

taxonID	scientificName	acceptedNameUsageID	taxonomicStatus
1	Coeligena helianthea	1	accepted 合法名
2	Ornismya helianthea	1	synonym 同物異名
3	Helianthea helianthea	1	synonym 同物異名

References.txt 資料文件

taxonID	Bibliographic citation
2	Schmidt, O. 1870. Grundzüge einer Spongien-Fauna des atlantischen Gebietes. (Wilhelm Engelmann: Leipzig): iii-iv, 1-88, pls I-VI
2	Laubenfels, M.W. De 1942. Porifera from Greenland and Baffinland collected by Capt. Robert A. Bartlett. Journal of the Washington Academy of Sciences 32(9): 263-269.

Other Synonymy Do's and Don'ts 其他同種異名該做與不該做的地方

- *dwc:acceptedNameUsageID* 必須指向資料集中現有的記錄，指向一個不存在資料集中的合法分類群是無效的。
- 不要混淆了用來描述分類的 *dwc:higherTaxonID*，與用來描述每筆學名紀錄分類狀態的 *dwc:acceptedNameUsageID*
- 同物異名之間不要相互「鏈結」。同物異名應該只藉由 *dwc:acceptedNameUsageID* 字串指向正式認定的分類階層記錄，而永遠不應該指向另一個同物異名。

⁷ http://rs.gbif.org/vocabulary/gbif/nomenclatural_status.xml

Nomenclatural Synonymy 命名上的同物異名

在核心資料檔案中，可以透過使用 *dwc:originalNameUsageID* 欄位來記錄同物異名資料的命名。這個字串指向的行列，為學名最初的分類依據。這個記錄建議透過 *dwc:namePublishedIn* 欄位來提供引用文獻書目，以表示最早發表該名稱的出版物。

taxonID	scientificName	originalNameID	namePublishedIn
1	Tetrao afer Müller 1778	1	J. Syst. Nat 7:31
2	Pternistes afer (Müller 1778)	1	
3	Francolinus afer afer (Müller 1778)	1	

命名上和分類上的同種異名可以被指定在同一個分類階層的記錄。

附註：*dwc:originalNameUsageID* 必須指向資料集中現有的記錄，指向一個不存在資料集中的合法分類群是無效的。

Pro-parte Synonymy 部分同種異名

有時一個名字可能會是多個合法學名的同種異名，或本身既是一個正式認定的合法學名又是另一種名的同物異名。這些都可能因為分類觀點的歧異，造成學名的處理的分合，例如模式標本可能被劃分成不同分類群。在下面的表格裡說明了分享部分同種異名的建議操作方法。在這個案例中，*Vireo solitarius* 本身是一個已正式認定的學名，它同時也是 *Vireo cassinii* 和 *Vireo plumbeus* 這兩個物種的同種異名。在同種異名的案例裡，個別以單獨一筆記錄呈現，在 *dwc:acceptedNameUsageID* 字串裡與正式認定的分類階元鏈結，彼此用("|")的符號隔開。

taxonID	scientificName	acceptedNameUsageID	taxonomicStatus
1	Vireo solitarius	1	accepted 合法名
2	Vireo cassinii	2	accepted 合法名
3	Vireo plumbeus	3	accepted 合法名
4	Vireo solitarius	2 3	Pro-parte 部分異名

Citation and Attribution 引用和歸屬

分類學名錄往往代表著編譯者個人和機構所付出的智力和財力。有些名錄可能是引用或是參考其他的名錄資源，來創造新的專題性、區域性或是分類學的觀點。因此，仔細說明資料的歸屬，以增加這些來源的能見度和貢獻，是必要且重要的。

DwC-A 格式對於適當的引用方式及所有權歸屬，提供了一系列的選項及建議。此範圍上至部分後設資料資源的全球引用和歸屬，下至記錄層級（record-level）資料元素。這些選項足以完整記錄個別資料來源的歸屬(multiple levels of attribution)。

Metadata Citation and Attribution 後設資料引用和歸屬

GBIF 的後設資料概況（profile）支援資源級別（resource-level）資料元素，有助於引用和歸屬，並詳細描述名錄的範圍和出處。對於所有後設資料元素的完整參考列表，已超出本文件討論的範圍，但是可在此網址⁸取得。具體引用和歸屬相關的資料包括：

- **Intellectual Property Rights 智慧財產權** – 後設資料概況（profile）包含對於資源（resource）的權利管理聲明，或是提供如創用 CC 授權條款的參考資料等，並且還包括描述資料集的可能用途和目標。
- **Individual and Organizations 個人和組織** – 後設資料配置文件允許任何一個與資料集相關的個人、機構或組織的說明描述。這些單位可能在資料集中個別扮演相關的角色，並包括每個資料的原始網址。本節對於物種名錄有貢獻的個人和組織，提供了描述和連結的方法。
- **Source URL 資源網址** – 連接到資源的首頁。
- **Project Information 計畫項目資訊** – 如果名錄連結到特定的計畫項目（例如：「全球物種名錄」The Catalogue of Life），則有一個欄位能詳細描述計畫項目的內容。
- **Citation 引用** – 此部分允許名錄發布者明確的規定物種名錄資料被使用時要如何引用來源。例如“Appeltans W, Bouchet P, Boxshall GA, Fauchald K, Gordon DP, Hoeksema BW, Poore GCB, van Soest RWM, Stöhr S, Walter TC, Costello MJ. (eds) (2011). World Register of Marine Species.” 網址為 <http://www.marinespecies.org> on 2011-02-22.”
- **Bibliography 書目** – 在後設資料文件可描述和涵蓋完整的參考書目來源。

⁸ http://links.gbif.org/gbif_metadata_profile_guide_en_v1

Data-level Citation and Attribution 資料層級的引用和歸屬

在每個資料集中，將引用和歸屬資訊紀錄在後設資料的文件中是很常見的。在某些情況下，甚至仔細到每一筆記錄。為了因應這些要求，下面列舉幾個記錄層級的語法（record-level terms），以詳述引用和歸屬的資訊。

- *dwc:nameAccordingTo*：記錄中可用此欄位來標識個人，或是引用的分類學參考文獻。（例如 “Erpenbeck, D.; Van Soest, R.W. M. 2002. Family Halichondriidae Gray, 1867. Pp. 787-816. In Hooper, J. N. A. & Van Soest, R. W. M. (ed.) Systema Porifera. A guide to the classification of sponges.”）
- *dwc:nameAccordingToID*：可回傳上述 *dwc:nameAccordingTo* 參考文獻的獨特識別碼。例如一個網址的連結。
- *dwc:datasetName*：如果記錄是從外部的資料集所獲得，則這個資料集可用文字串的形式來引用。（例如 “World Register of Marine Species, cited on 12 April 2011”）
- *dwc:datasetID* -代表資料集的識別碼，最好是可辨認的。
- *dc:source* -連結到資料網頁來源。

Authenticated File Access via httpd 透過 httpd 身分驗證存取檔案

GBIF 主張自由且公開地存取生物多樣性資料，並且擴展至分類學資訊的範圍。儘管如此，一些分類學資料發佈者，在沒有跟資料使用者做任何協商下的情況下，可能會對於在完全公開的公用網址上提供資料感到猶豫，即使資料本身是免費且公開。對於這樣的情況下，可以透過達爾文核心集檔案的發布，經由通過認證的用戶名稱和密碼登入後才存取資料。網址本身可以自由發布，但是存取檔案則需要通過身分驗證。資料發布者可以使用網路伺服器日誌記錄功能，經由特定的使用者來追蹤存取的情形，但是，值得再次重申的是，這並不是 GBIF 偏好的方法。GBIF 的立場是，若有分類學資源的需求，GBIF 偏好具一致性、對使用者友善的引用和歸屬的過程，正如在本文裡所定義的。

Use Case #1 – Checklist composed of multiple contribution datasets (e.g., Catalogue of Life, PESI, WoRMS)

使用案例 #1 - 由多個資料集組成的名錄（例如：Catalogue of Life, PESI, WoRMS）

分類學資料集可能是多個資料來源的組合，除了資源的集體來源本身，其中每一筆來源都必須被認可（acknowledged）。這樣的例子有很多。或許最大的集體努力全球物種名錄一年度名錄（Catalogue of Life Annual Checklist），致力於提供完整的世界生物名錄。物種名錄本身是由各個獨立的主分類群資料集所組成，每一個資源可能是由專門的子網路專家所貢獻。

其他例子包括「泛歐洲物種名錄」(Pen-European Species list)，它由歐洲動物相(Fauna Europaea)、歐洲海洋生物物種登錄系統(European Register of Marine Species)、歐洲+地中海植物資料庫(Euro+Med PlantBase)及其他資料集所組成。另一個例子則是「全球海洋生物物種登錄系統」(World Register of Marine Species)。

為了有效地記錄各式各樣不同資源的來源，建議做法歸納如下：

1. 建立一個單一的後設資料文件，來代表集體資源本身（例如 The Catalogue of Life、The World Register of Marine Species 等等），後設資料文件提供了適當的引用、代理人、權利，和前文所述的其他元素。文件檔名可註記為被引用做達爾文核心集檔案描述者文件 meta.xml，以便直接與整個達爾文核心集檔案資料集連結在一起。最佳的建議做法是使用 GBIF 後設資料配置方式，並將文件命名為 EML.xml。在這種情況下，後設資料描述者 XML 將會看起來像這樣：

a. `<archive xmlns="http://rs.tdwg.org/dwc/text/" metadata="eml.xml">`

2. 每個組成元件的資料集，皆可建立額外的後設資料文件，並加以歸檔。這使得每個子組成資料集跟親資料集一樣，可以完整地記錄下它們自己的建議引用文獻、貢獻者等等。由於這些資料集沒有記錄所收集的全部資料，所以不用 meta.xml 來描述這些檔案。取而代之地，可以透過 `dwc:datasetID`，透過個人資料記錄來加以引用。如果歸檔文件本身就包含後設資料文件，則 `dwc:datasetID` 等同於該文件的檔名。另外，它可能指向一個 URL，或是一些獨特和具有解析力的識別碼，以提供資訊的了解。一個較不推薦但是可行的替代方法是，將 URL 添加到一個描述資料集的簡單網頁，而不是一個結構化的後設資料文件。
3. 將個人資料引用於記錄層級(record level)時，也就是第 3 層級的引用，建議使用 `dwc:nameAccordingTo` 欄位。額外的記錄層級語法已於上述文章中提供。

Use Case #2 – Checklist derived from one or more authority sources

使用案例 #2 – 從一個或多個授權來源獲得的物種名錄

在本案例中，物種名錄是為了特定的目標而編輯，但是它的基本分類結構來自一個或多個外部分類學名錄，可視為「權威檔案」(authority files)。新的編纂可包含對基本資料記錄的額外註記，以便符合新的名錄著重的目標。舉例而言，歐洲的國家物種名錄是來自於如歐洲動物相 (Fauna Europaea)、或全球物種目錄(The Catalogue of Life)的資料集，在原則上，它們(譯註：它們應該是指 Fauna Europaea 或是 The Catalogue of Life)可以提供一個國家的完整名錄，做為它們自己資料庫內的一個子集。國家名錄中可以再增加額外的區域性紀錄或指標，例如該地區對物種受威脅狀態的評估或是一些感興趣的欄位等，而成為一個新的、衍生的資料集。在這種情況下，能夠提供記錄層級的歸屬、以及資料集來源的連結是非常重要的。達成的建議方法如下：

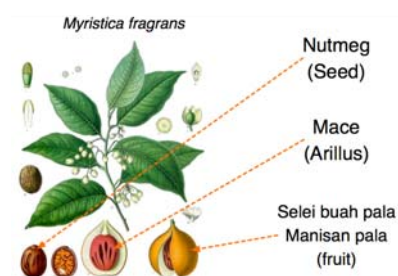
1. 建立一個單獨的後設資料文件來代表新的、衍生的資源本身(例如，荷蘭國家物種名錄 National Checklist of the Netherlands)。在此後設資料文件中，所舉出的資料集是可以被引用的。
 - A. 充分說明組織的參與者角色，並與資源網站做連結。
 - B. 在參考書目資料章節中引用，呈現方式如參考資料集所建議。
2. 在數據文檔裡，可以在記錄階層裡做額外的歸屬和連結。這包括：
 - a. 在 `dwc:datasetName` 裡，依名字引用資料集。
 - b. 在 `dwc:datasetID` 裡，依 ID (例如 URL) 引用資料集，並連結到資料集首頁。
 - c. 利用 `dc:source`，提供連結到引用資料集網頁裡相對應物種的頁面。
 - i. 如果 `dc:source` 被保留為指向衍生資料庫的 URL，仍然可以利用替代辨識碼延伸(Alternative Identifiers extension) 連結到資源資料庫。
 - d. 如果資源資料集提供物種名錄中引用分類單元的全球唯一識別碼 (globally unique identifiers)，它們可以當作衍生的資料集中的 `taxonID`。這可以確保清楚明確地聯繫到原始的分類單元，如果可能的話，強烈建議使用。

- e. 使用 *dwc:nameAccordingTo* 或是 *dwc:nameAccordingToID* 來指向對應的原始記錄中的分類單元定義，像是引用文獻或是 URL。

Sharing Vernacular Names 共享俗名

全球名稱體系概況(GNA profile)支援物種名錄中與分類階層有關的俗名資料共享，其作法是用「俗名延伸法」(Vernacular Names extension)⁹把俗名當做獨立的、關連的檔案來分享。這種延伸提供了一套豐富的俗名描述用詞，包括地區性的和型態的修飾語等等。完整的延伸條款和建議的詞彙列表，可以在先前註解的 GBIF Resource Repository 中，或是從全球名稱體系概況引用指南(GNA Profile Reference Guide)¹⁰裡查詢。

圖： *Myristica fragrans* 肉豆蔻、Nutmeg (seed) 肉豆蔻種子中的核仁（種子）、Mace (Arillus) 肉豆蔻衣（假種皮）、Selai buah pala Manisan pala (fruit)（譯註：用當地語言表達“肉豆蔻的果肉”）（果實、水果）



俗名可透過擴充檔來引用，因此它們必須連結到親階核心資料檔案裡已命名的分類階層。更進一步的建議是，俗名記錄提供語言的參考，藉由俗名的使用來辨識俗名使用的語言。最好的作法是使用 ISO 639 語言代碼來分享語言資訊。整套完整的語言代碼可以參考 GBIF vocabulary server¹¹。俗名也可能有獨特的地區性用法，可以透過用 *dwc:locality* 詳述，或是粗略地使用 *dwc:country* 欄位。建議使用 ISO 6133 國家代碼來表示國家名稱，這也可以在 GBIF vocabulary server¹²上查詢。

Sharing Species Descriptions 共享物種描述

全球名稱體系概況(GNA profile)可透過 Taxon Description extension¹³支援分類階層相關的描述性資訊分享。可以指定描述性資料到不同的描述類型中，而且，在擴充檔中發布資料時，多個描述性記錄可以被連結到單一個分類階層，讓每個分類階層擁有相對豐富的資料。建議使用 descriptive type vocabulary¹⁴來詳述描述性資訊。

⁹ Vernacular Names Extension <http://re.gbif.org/extension/gbif/1.0/vernacularname.xml>

¹⁰ http://links.gbif.org/gbif_gna_profile_reference_guide

¹¹ <http://vocabularies.gbif.org/vocabularies/lang>

¹² <http://vocabularies.gbif.org/vocabularies/country>

¹³ <http://rs.gbif.org/extension/gbif/1.0/description.xml>

Multi-line descriptions 多段描述

描述性資訊應該限制在同一段文字段落區塊內。應該要避免或小心輸入含有分行符號的多個段落，以維持做為達爾文核心集檔案合成的文字檔輸出的完整性。多行資料欄位作為文字檔案時，需要記錄界定符號，通常是用分行符號，來區別多行資料欄。支援單一欄位多行的最好方法，是將分隔字元以非分隔字元或是字元組取代，因此當資料被利用與解析時，使用者可以用適當的分隔取代。其中一個例子是以 html 語言的
 分行標籤取代一般的分行符號。

Sharing Species Distributions 共享物種分布

全球名稱體系概況(GNA profile)透過 Species Distribution extension¹⁵ 支援物種分布資料的分享，以便能發布每一個分類群的多個描述記錄。物種分布的擴充檔不僅僅被用來指示國家或區域的分布描述，同時也支援各分類群受威脅狀況的引用分布的資格性，例如該物種是否為被引進種、或是原生種等等，以及可能與特定區域綁在一起的其他屬性。

指定一個特殊區域的最佳建議操作方式是經由 dwc:localityID 欄位，透過可辨識的或是眾所週知的區域識別碼來發布。

如果使用 dwc:country 欄位，建議使用 ISO 3166 國碼，該資訊在 GBIF vocabulary server¹⁶ 網頁上可取得。

Sharing References 共享參考文獻

全球名稱體系概況(GNA profile)經由 Reference extension¹⁷ 支援參考文獻書目分享。建議使用 Reference extension，它是被設計用來分享在專論與已註記物種名錄裡的同物異名資訊。它能分享分段式引用資料，因此提供更詳細的引用格式，如在核心資料檔案中的一些引用-儲存資料元素，像是 dwc:namePublishedIn。

¹⁴ http://rs.gbif.org/vocabulary/gbif/description_type.xml

¹⁵ <http://rs.gbif.org/extension/gbif/1.0/distribution.xml>

¹⁶ <http://vocabularies.gbif.org/vocabularies/country>

¹⁷ <http://rs.gbif.org/extension/1.0/reference.xml>

這個擴充檔經由 *dc:type* 欄位支援資料分類學和命名學的合格性的參考，當與 the Reference Type vocabulary¹⁸ 一起使用時，可以被用來辨別與分類單元相關的一組參考文獻。

Sharing Type information 共享模式標本資訊

全球名稱體系概況(GNA profile)藉由模式標本及標本的擴充檔 (Types and Specimens extensions¹⁹) 來支援有關模式標本與標本的資訊共享，包括模式標本、模式種與屬等基本資料共享。

Sharing Links and Identifiers 共享連結及識別碼

全球名稱體系概況(GNA profile)支援共享的方式和描述多個相關外部資源連結的工具²⁰。它藉由可辨識的識別碼 (resolve-able identifiers) 讓資料發布者與原始資料庫或文件連結。提供給單獨一個分類單元的多個辨識碼，可連結到網頁或是機器可判讀的網路服務回應 (a more machine-readable web service response)。在識別碼是可以辨識的 (resolve-able) 情況下，建議每一筆資料的格式都能讓使用者瞭解該如何闡述回應的資訊，這通常是以 mime type 來達成。完整的 mime types 清單可以前往此網址²¹ 取得。

Creating a dynamic link to species page 創造物種頁面的動態連結

連接回原始資料庫的連結，通常都遵循共通的格式，僅在辨識碼的數字或網址上使用的分類單元名稱有所不同。但這會導致冗長且膨大的檔案名稱。達爾文核心集檔案格式支援更有效率的定義網址模版 (URL template) 的方法，只需要定義一次，而且允許變數嵌入模版中，以消除重覆需要資料檔案中每個分類單元的網址的情形。以上方法不需要使用參考資料擴充 (Reference extension)，可以透過達爾文核心集檔案的 XML 後設檔案 (metafile) 組件達成，這需要對 XML 有一定程度的熟悉來編輯 XML 後設檔案。GBIF 提供達爾文核心後設檔案 (Darwin Core metafile) 的詳盡指南²²。

¹⁸ http://rs.gbif.org/vocabulary/gbif/reference_type.xml

¹⁹ <http://rs.gbif.org/extension/gbif/1.0/typesandspecimen.xml>

²⁰ <http://rs.gbif.org/extension/gbif/1.0/identifier.xml>

²¹ <http://www.iana.org/assignments/media-types/index.html>

²² http://links.gbif.org/gbif_dwc-a_metafile_en_v1

後設檔案支援在後設檔案中創造變數，可以是一個網頁或是網路服務要求（web service call）。這個變數可以嵌入網址，並包含一個分類單元識別碼，或是以分類名稱為參數的網址。已發布資料的任何一欄皆可引用，在大括弧“{}”裡放入指數即可。在核心資料檔案裡的分類單元辨識碼，也可以經由變數“{id}”來引用。以下範例說明這些特徵：

1. 整合分類資訊系統（Integrated Taxonomic Information System，ITIS）利用分類學序號（Taxonomic Serial Numbers，TSN）提供網站內各個分類單元頁面的連結。

http://www.itis.gov/servlet/SingleRpt/SingleRpt?search_topic=TSN&search_value=174375

如果核心資料檔案以ITIS TSN系統發布，則可以形成一個連結，並藉由以下語法（syntax）與核心資料標準裡的辨識碼綁定。

以記錄的id為基礎連結到物種頁面：

```
<field
```

```
default="http://www.itis.gov/servlet/SingleRpt/SingleRpt?search_topic=TSN&search_value={id}"
```

```
term="http://purl.org/dc/terms/identifier"/>
```

上述語法中，原本的數值被變數“{id}”取代。這數值是從核心ID獲得。

2. 2010年年度全球物種名錄（2010 Catalogue of Life Annual Checklist）提供相似的識別碼，同時也能支援以名字為基礎的搜尋，並能嵌入在網址裡。例如：嵌入物種學名“Struthio camelus”到網址裡。

<http://www.catalogueoflife.org/annual-checklist/2010/search/all/key/Struthio+camelus/match/1>

利用達爾文核心集語法“scientificName”，可以在核心資料檔案中發布完整物種學名的組合。假設scientificName呈現在核心資料檔案中的第12欄，我們可以使用以下語法：

以記錄的id為基礎連結到物種頁面：

```
<field default="http://www.catalogueoflife.org/annualchecklist/
```

```
2010/search/all/key/{12}/match/1"
```

```
term="http://purl.org/dc/terms/identifier"/>
```

上述語法中{12}代表在網址裡將要被取代的第12欄位的數值。