

# 生物類數位典藏後設資料 (METADATA)建置與整合之研究

---

陳淑君 陳亞寧

中央研究院計算中心

數位典藏國家型科技計畫  
後設資料工作組

2002年11月27日

# 大綱

---

- 研究背景，研究方法
- 文獻探討
  - 國際間生物類典藏品後設資料的發展
    - 標準與類別
    - 後設資料與大環境關係、發展與定位
  - 結果與發現
- 個案研究
  - 數位典藏國家型科技計畫個案
    - 屬性分析
    - 後設資料建置與整合狀況
  - 結果與發現
- 結語：研究成果與未來研究建議

# 研究背景 & 研究方法

- 隨著自然史數位典藏作業的展開，如何建構一個可以再利用與共享的資訊環境，已成為國際間共同的焦點。
- 本研究主題在於探討與分析目前國際間處理生物類典藏品後設資料的方法與特色，並釐定幾套重要的後設資料國際標準。同時，以數位典藏國家型科技計畫為個案觀察所屬相關計畫如何發展後設資料標準，以及彼此間的整合方式。
- 文獻探討法，個案研究法

# 生物類資料標準與類型

- 從國際生物科學聯盟 – 分類資料庫工作小組 (TDWG) (International Union of Biological Sciences, Taxonomic Database Working Group)的視野，將標準分成三大類型：
  - 語義性／資料結構定義標準 (semantic standards / data structure definition)
  - 語法性／資料交換標準 (data { exchange | interchange | transfer } standards)
  - 字彙性／權威控制標準 (vocabulary or authority standards)

# 生物類資料標準與類型：

## 語義性／資料結構定義標準

- 規範資料元素的定義或更複雜的資料結構
- 範例：
  - Plant Names in Botanical Databases (Bisby 1994) <http://www.tdwg.org/plants.html>
  - ASC (Association of Systematics Collections) Reference Model for Biological Collections (ASC 1993)
  - UCMP (University of California Museum of Paleontology) Data Model (UCMP 1994)
  - DarwinCore (DwC) (Species Analyst 2001)
  - FGDC Biological Metadata Profile (FGDC 1999)

# 生物類資料標準與類型：

## 語法性／資料交換標準

- 規範資料流 (data stream, ex. File) 語法與語義的組成，使資料能在系統間彼此“複製”
- 範例：
  - HISPID 3 (Conn 1996)
  - XDF(EXchange Data Format) (White & Allkin 1992)
  - ITF (International Transfer Format for Botanic Garden Records)(Jackson 1997)
  - Data Specification, Common data exchange format - Core Data (Flora of Texas Consortium 1997)
  - BarCode Technology for Herbarium and Museum Biological Collections (Conn 1997)

# 生物類資料標準與類型：

## 字彙性／權威控制標準

- 規範整套的資料值或記錄，使資料能在不同系統間直接可以匹對(**comparable**)
- 範例：
  - POSS (Plant Occurrence & Status Scheme)  
<http://rbge-sun1.rbge.org.uk/BG-BASE/poss.htm> (Leon et al., 1995)
  - World Geographical Scheme for Recording Plant Distributions (Hollis & Brummit 1992)
  - Authors of Plant Names (Brummit & Powell 1992)

# 後設資料與大環境關係、發展及定位

- 重要會議
  - Metadiversity I, II, III (1998)(2002)(2003)
  - Trends and Developments in Biodiversity Informatics Forum (Oct. 2002)
- 重要組織與計畫
  - 國際性: GBIF, CHM, TDWG, Species 2000
  - 洲際性: IABIN, BioCASE
  - 國家性: NBN, NBII



# GBIF (The Global Biodiversity Information Facility)

- 全球生物多樣性資訊中心 (<http://www.gbif.org/>)
- 後設資料的角色與定位
  - 建立資料標準與交換機制，以整合物種及標本層次的資料至**GBIF**網路。
  - 物種名稱權威控制，典藏資料，整合與互通等三方面皆設置專門委員會研究相關議題與標準
  - Sub Committees: Data Access & Database Interoperability (DADI)
- 2003年預計推薦的後設資料相關標準
  - 跨資料庫檢索協訂：**DiGIR** Protocol (**D**istributed **G**eneric **I**nformation **R**etrieval)
  - 典藏資料交換規格：**ABCD** Schema (The **A**ccess to **B**iological **C**ollections **D**ata)

# CHM (Clearing-House Mechanism of the Convention on Biological Diversity )

- 生物多樣性資訊交換機制  
(<http://www.biodiv.org/chm/>)
- 主要功能在於倡導保育工作相關科學與技術上的合作，並俾便生物多樣性資訊的檢索與交換。
- 後設資料的角色與定位:作為**CHM**基本架構，讓相關資訊的載入或轉出皆遵循一定的格式
- 目前推薦採用的後設資料標準
  - Z39.50, Dublin Core, GILS (Government Information Locator Service), FGDC Biological Data Profile (Federal Geographic Data Committee)

# TDWG (Taxonomic Database Working Group)

- 分類資料庫工作小組  
(<http://www.tdwg.org/index.html>)
- 後設資料的角色與定位
  - CODATA (Committee on Data for Science and Technology, 國際科技資料委員會), GBIF, Species2000...
- 目前公佈採用的後設資料標準 (15個)
  - Authors of plant names
  - ITF (International Transfer Format for Botanic Garden Records)
  - Botanico-periodicum-huntianum
  - HISPID3...

# IABIN (Inter-American Biodiversity Information Network)

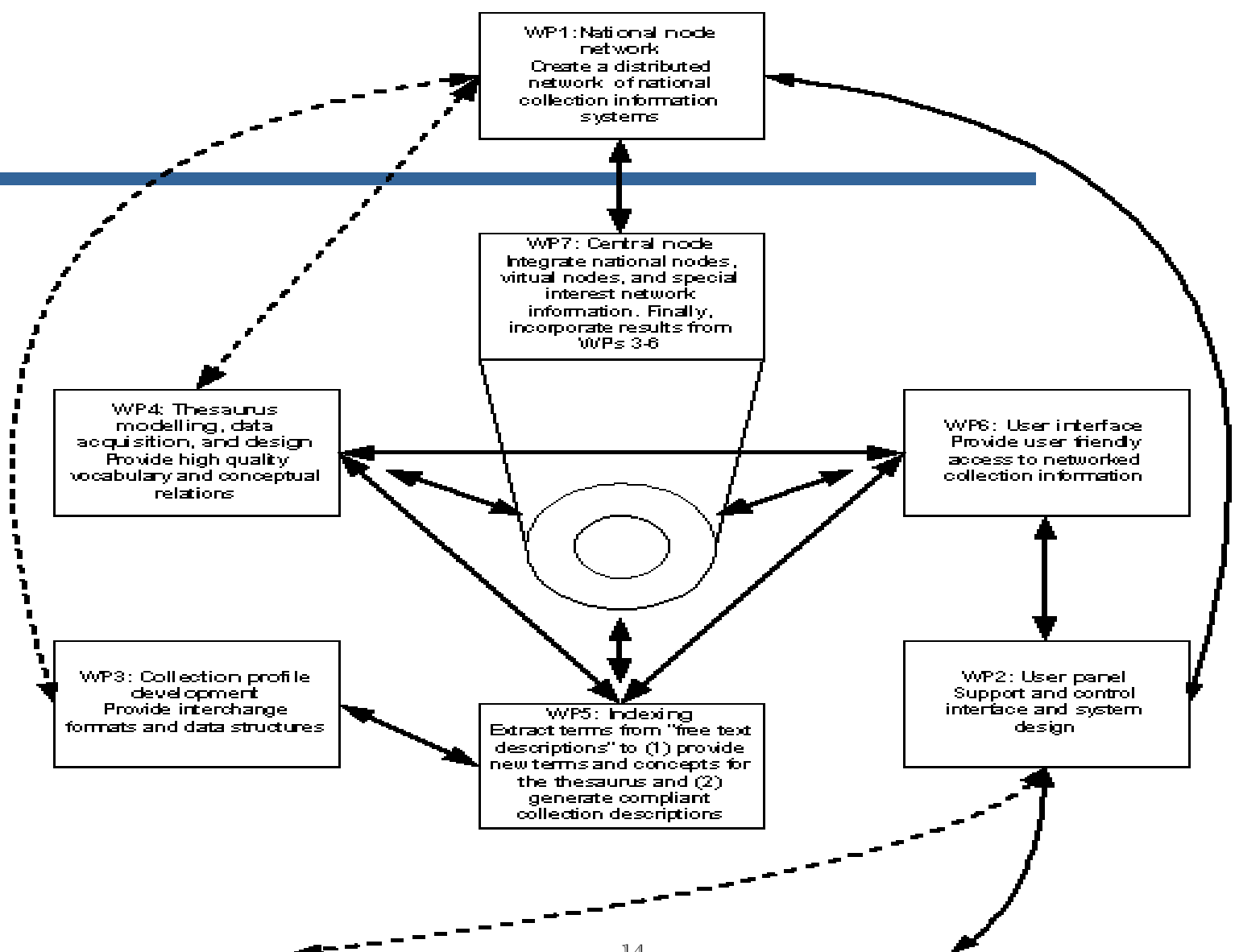
---

- 美洲間生物多樣性資訊網路  
(<http://www.iabin-us.org/>)
- 跨國跨政府的活動計畫, 旨在倡導泛美洲地區的國家, 能更有效地收集、共享與使用生物多樣性相關資訊
- Survey of New World Holdings in Museum and Private Collections in Europe (1999)
- Harmonizing Metadata Initiatives Throughout IABIN (1999)

# BioCASE (A Biological Collection Access Service for Europe)

- 歐洲生物典藏品檢索服務  
(<http://www.biocase.org/>)
- 後設資料的角色與定位
  - Workpackage 1-7
    - WP3 - Collection Profile Development : 提供交換格式與資料結構
    - WP4 – Thesaurus modelling : 提供高品質的字彙控制與概念關聯性的建立
- 目前公佈的後設資料相關標準
  - ABCD Schema (The Access to Biological Collections Data)(Nov. 7, 2002)
  - BioCASE Thesaurus
  - BioCISE Model

# BioCASE工作組主要合作關連圖 (WP1-7)



WP8: Legal basis and best practice      14      WP9: Business plan and sustainability

Build a solid basis and a business model to ensure a sustainable service after the project's end; help to resolve IPR issues affecting the collaboration of collections, national nodes, central node, and special interest networks

# NBII (National Biological Information Infrastructure)

- 美國國家生物資訊基礎建設 (<http://www.nbii.gov/>)
- 後設資料的角色與定位
  - 扮演橋樑的角色, 讓生物科學的社群與其他相關社群間 (如: 地理資訊社群) 的資訊基礎建設得以連結與互通.
- 目前公佈的後設資料標準
  - **Biological Data Profile**

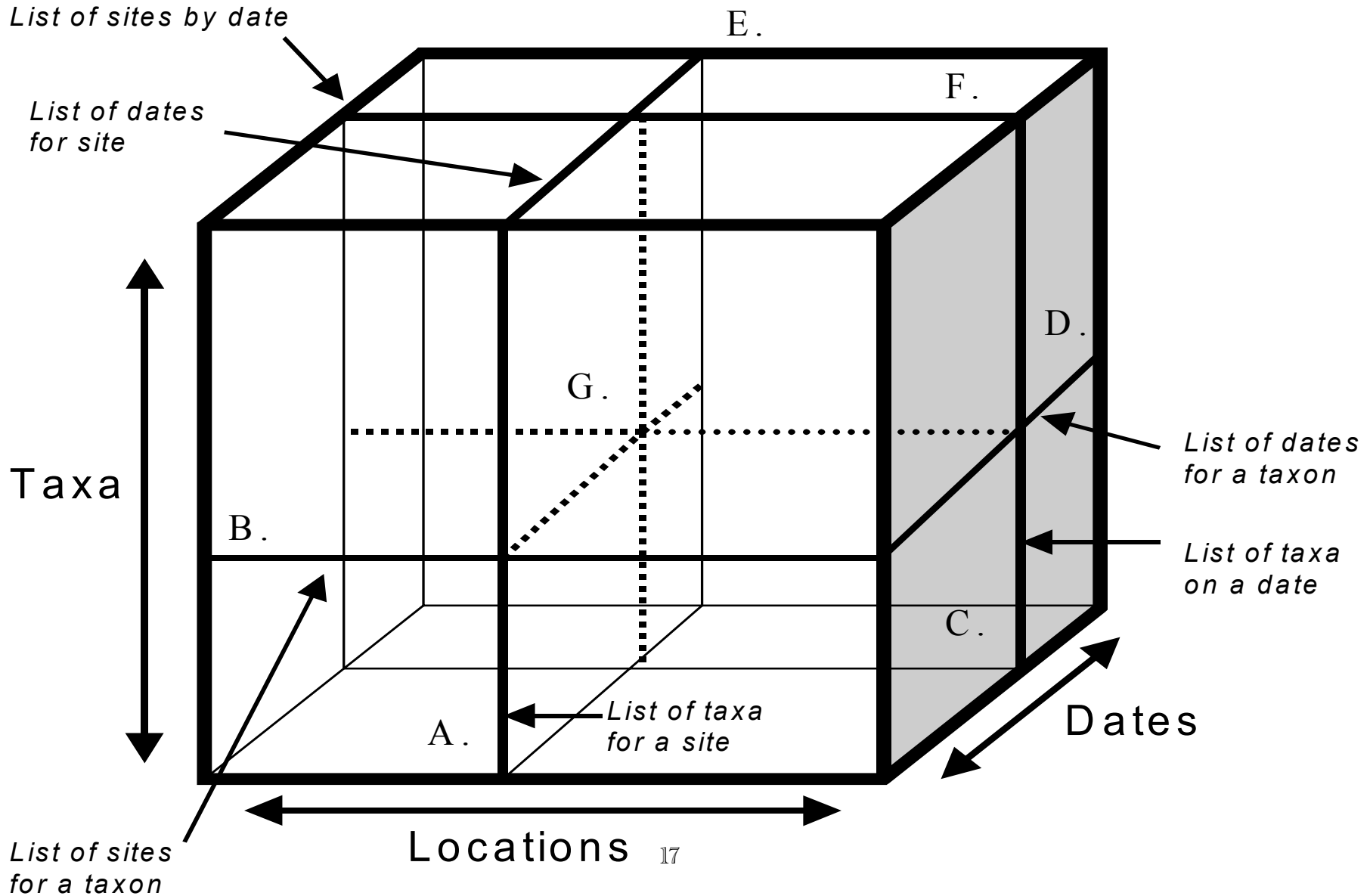
# NBN (National Biodiversity Network)

---

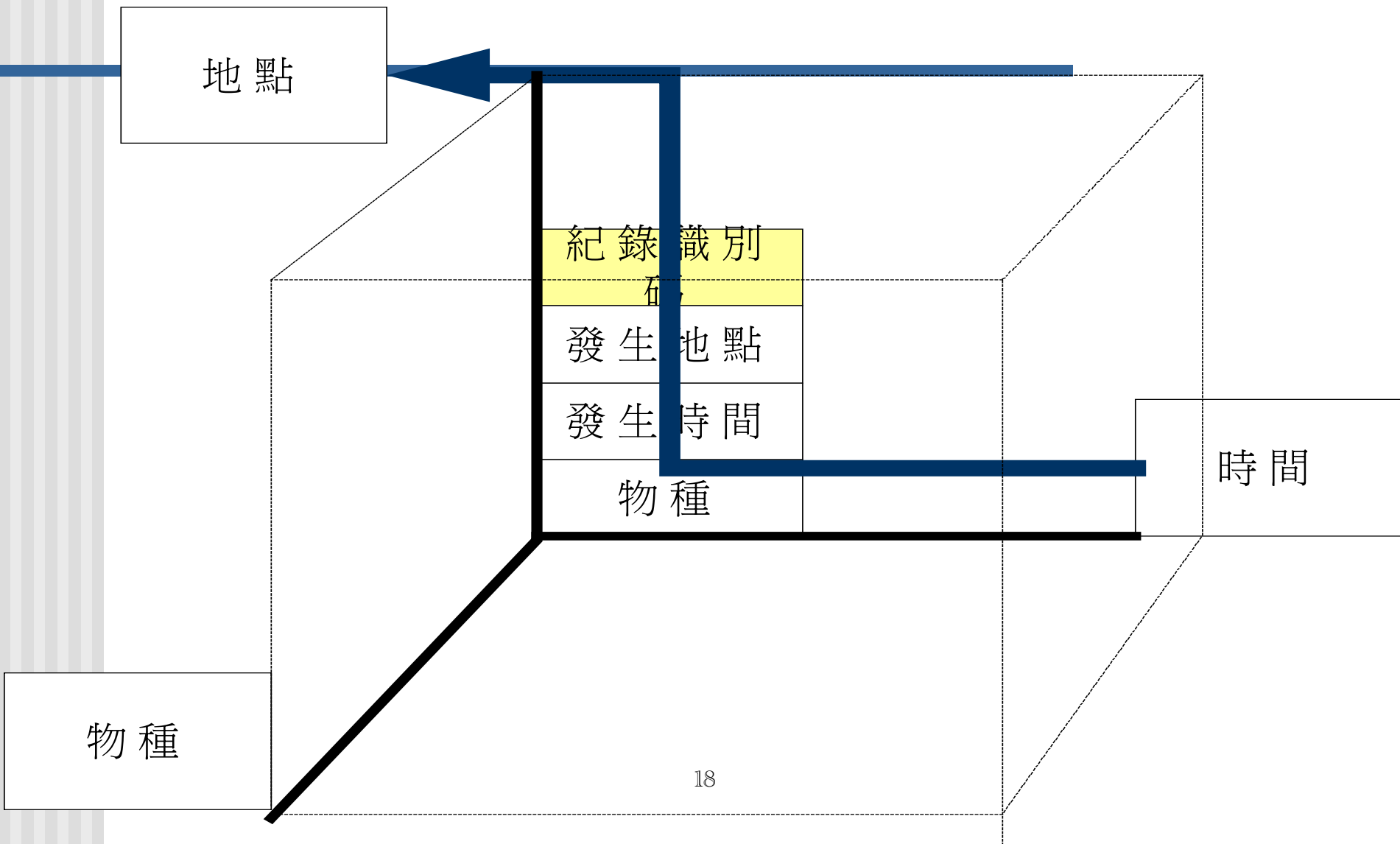
- 英國生物多樣性網路  
(<http://www.nbn.org.uk/>)
- 後設資料的角色與定位
  - Data Standard Project: 包括發展與確認 Metadata Standard、Data Models、Data Exchange、Standard Dictionaries、Standard Access terms and conditions等內容.
- 目前公佈的後設資料標準
  - NBN Data Model



# Biological Records Cube (生物紀錄角錐)



# 「紀錄」「地點」「物種」「時間」立體角錐 呈現之關聯圖



# 文獻探討結果與發現<sup>1</sup>

- 後設資料的重要性，已由傳統資料庫與圖書館發展時的隱性與附屬角色，成為數位典藏大環境的核心與中堅的角色
- 後設資料的發展重點
  - 互通性（Interoperability）
  - 標準化（Standardization）
- 生物類後設資料，著重於發展與遵循領域導向(domain-specific)的資料標準

# 文獻探討結果與發現<sup>2</sup>

- 領域導向的後設資料發展，由早期的單一子領域（如：植物，昆蟲...），至目前的整體生物學領域的整合性發展（如：BioCISE提出的 A comprehensive reference model for biological collections and survey）
- 後設資料標準的採用呈現多樣性
  - 不同標準應用於不同層面（如：分類，標本，文獻，分佈地區...）
  - 同一層面也應用不同標準（如：多套的資料模式，綱要，協定，交換標準...）

# 文獻探討結果與發現<sup>3</sup>

- 後設資料標準的未來發展趨勢
  - 將由GBIF + TDWG + BioCASE + Species 2000...逐漸達成標準與互通機制方面的共識
- 後設資料的處理與含蓋範圍
  - 將會由著重標本與物種層次的描述，延展至生物多樣性(Biodiversity) 所有層次生物組織，包括：基因(genetic)、物種(species)與生態系(ecosystem)三種層次，以及建置與整合彼此間的關聯性

# 個案研究：數位典藏國家型科技計畫－屬性分析<sup>1</sup>

## ■ 11個計畫

- 台灣動物相典藏之研究：台灣魚類相之數位典藏（中研院）
- 台灣動物相典藏之研究：台灣貝類相之數位典藏（中研院）
- 動物模式標本典藏數位化子計畫（科博館）
- 台灣特有及重要館藏標本典藏數位化（科博館）
- 動物博物館典藏數位化計畫（台大）
- 昆蟲標本館典藏數位化子計畫（台大）
- 臺灣維管束植物相典藏數位計畫（科博館）
- 臺灣非維管束植物暨地衣子計畫（科博館）
- 台灣真菌之數位典藏（科博館）
- 植物標本館數位典藏計畫（台大）
- 台灣本土植物數位典藏（<sup>22</sup>中研院）

# 個案研究：數位典藏國家型科技計畫－屬性分析<sup>2</sup>

## ■ 機構與社群

- 台大（動物博物館，昆蟲標本館，植物標本館）
- 中研院（動物所，植物所標本館）
- 科博館（動物學組，植物學組）

## ■ 學科

- 魚，貝，昆蟲，維管束植物，非維管束植物，地衣，真菌...

## ■ 資料類型

- 標本照片，標本記錄，標本標籤，文獻紀錄，物種基本解說資料（生態照片，生態影片），聲音，模式標本...

# 個案研究：數位典藏國家型科技計畫—後設資料建置與整合狀況<sup>1</sup>

- 內容分項計畫 (動物小組, 植物小組)與數位典藏國家型科技計畫後設資料工作組
- 整合重點：「後設資料共通元素」的釐定與彙整
- 重要里程碑與進度成果
  - 共通元素
    - 2002/1/25 第一次動物暨植物主題小組會議：決議「動物組與植物組分別制定後設資料，再整合共通性後設資料」
    - 2002/1/25, 3/7, 3/26 植物學組召開3次會議，完成「植物學組後設資料總表」與「植物學組共通元素」
    - 2002/3/1, 4/3, 5/6 動物學組召開3次會議，完成「動物學組後設資料需求元素總表」與「動物學組共通元素」
    - 2002/6/17 第二次動物暨植物主題小組會議：完成「動植物小組後設資料共通元素比對表」



# 個案研究：數位典藏國家型科技計畫—後設資料建置與整合狀況<sup>2</sup>

## ■ 共通元素（續）

- 元素名稱（中文/英文），元素定義，資料值清單，著錄範例，相關國際標準比對
- 目前整合的共通元素（包括：物種名錄，物種生物資訊，標本資料與記錄，文獻資料，媒體資料與影像資料等類別）總計；30 個共通元素。（動物學組共通後設資料元素92個；植物動物學組共通後設資料元素42個）

## ■ 後續的問題研析

- **2002/7-8** 完成「標本採集地理資訊後設資料及行政區劃分調查與建議報告」

# 個案研究：數位典藏國家型科技計畫—後設資料建置與整合狀況<sup>3</sup>

## ■ 國際標準研析與接軌

- 2002/3 完成「生物典藏資料後設資料標準清單與簡介」
- 2002/3 完成「HISPID3 中譯初稿」
- 2002/3 完成「FGDC-Biological data profile中譯初稿」
- 2002/4 完成「Species2000簡介與物種名錄標準中譯」
- 2002/6 完成「地質學組後設資料研究報告」
- 2002/9 完成「台灣動物相典藏之研究：魚類後設資料需求規格書 0.2版」
- 2002/9 完成「台灣動物相典藏之研究：貝類後設資料需求規格書2.0版」
- 2002/9 完成「科博館資料庫後設資料完整性研究與建議報告」
- 2002/10 完成「臺灣本土植物數位典藏計畫後設資料需求規格書2.0版」

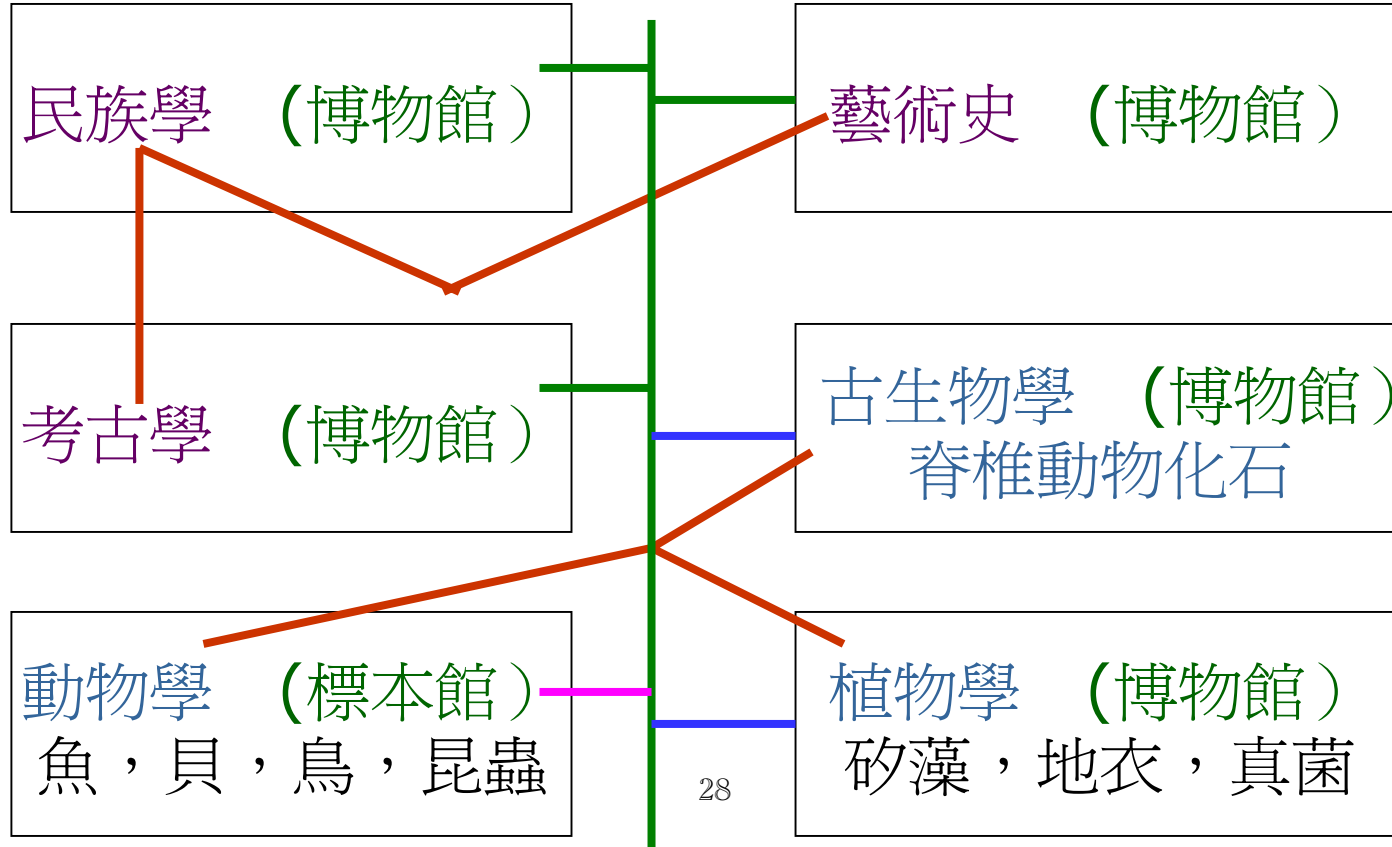
# 個案研究：結果與發現<sup>1</sup>

- 後設資料標準的應用
  - 規格書（Specification），綱要（Schema）
  - Spcies2000, HISPID3, DC (包括結合OAI的應用)
- 共通元素的發展
  - 溝通，共識，釐清問題/差異性/需求
  - 未來合作的基礎
  - 元素定義，資料值需再進一步釐清

主題小組	總需求元素	共通元素	共通元素
動物學	1 8 6 個	9 2 個	3 0 個
植物學	1 1 6 個	4 2 個	

# 個案研究：結果與發現<sup>2</sup>

## ■ 跨機構社群與跨學科領域的功能需求



# 結論與建議<sup>1</sup>

---

- 共通元素
  - 定義與功能
  - 定位與未來發展
- 後設資料作業的範疇
  - 擴展標準化的層面：語義，語法，字彙控制等規範與共識
  - 發展不同層級學科領域後設資料的規範與互通
  - 發展相同性質機構社群的核心元素

# 結論與建議<sup>2</sup>

- 後設資料國際標準的採用原則
  - 釐定共同需求：共通元素
  - 掌握生物類國際標準的主流
    - GBIF+ (學科領域：生物類)
    - CIDOC (機構社群：博物館)
    - 元素(element)，字彙(vocabulary)，綱要(schema)，協訂(protocol)...
  - 提出適用的「後設資料應用檔特徵」  
(Application Profile-Biological Collection)

# 誌謝

---

- 內容分項計畫
- 動物與植物主題小組
- 各主題計畫
- 後設資料工作組同仁