推動海峽兩岸生物多樣性資料庫之整合與分享

邵廣昭

中研院 生物多樣性研究中心

【摘要】海峽兩岸生物多樣性資料及資訊網之建置工作，在過去的七、八年內均各自已有很好的基礎及成果，特別在名錄、標本、分布、物種解說、影像、種原庫等以分類及生態爲主的方面，迄今已匯集甚多的數位化的資料。這些資料之建置，所採用之交換標準，海峽兩岸近年來也各自與 GBIF、COL（Species 2000）、BOL、EOL、ILTER、BHL、WoRMS、FishBase、OBIS、AquaMap 等不同的國際組織或合作計劃進行合作與交流。但遺憾的是海峽兩岸資料庫本身卻始終未能真正進行整合與實質上的合作。由於生物多樣性的研究、保育、教育及資源之利用與管理上均需仰賴完善及正確資料庫爲基礎，特別是兩岸生物多樣性地理親緣關係密切，物種之相似性高，更需要先推動兩岸資料庫之整合。故乃希望利用未來三年內透過兩岸互訪、網路、會議等不同方式，協商應採取之架構、格式標準與資料內容，逐步完成兩岸生物物種名錄，標本館典藏標本資訊（含生命條碼所需之酒精或液態氮組織標本），及科技名詞詞彙中譯名之對照與整合。共同建立「海峽兩岸生物多樣性資訊網」之網站，提供正體、簡體及英文三種文字，供外界使用。另外在參與國際組織之合作亦希望能攜手合作相互支持，在互信互惠之原則下創造雙贏。

【關鍵詞】生物多樣性資料庫、資料整合、兩岸合作、生態保育、數位典藏

一、 生物多樣性資訊學之發展

生物多樣性資訊學（Biodiversity informatics）是一門近年來才興起的學科，亦即在繼生物資訊學（Bioinformatics）整理分析龐大的遺傳及 DNA 序列資料之後，將更龐雜的物種及生態系的資料，包括從基因（含生命條碼）、形態、生態、生物學等與分類學、演化、生態、保育等基礎及應用科學所需的材料及資料，利用資訊科學及資料庫管理的技術予以蒐集、整合、分析，以掌握生物多樣性之變遷，並運用在預測及管理上，特別是未來自然或人爲因子衝擊下生物多樣性之改變，如氣候變遷及外來入侵種等，以確保全球生物多樣性不再減少，可採取更有效之因應措施，使環境及資源可以永續發展。在這門學科領域中已有多項國際合作計畫正如火如荼地展開，包括全球物種電子名錄建置計畫 Catalog of Life（COL）爲 Species 2000 及 IT IS 兩計畫的合作計畫，全球生物多樣性資訊機構（Global Biodiversity Information Facility, GBIF），生命之樹（Tree of Life, TOL），生命條碼 （Barcode of life, BOL），生命大百科（Encyclopedia of Life, EOL)，國際長期生態監測（ILTER），生物多樣性遺產文獻庫（Biodiversity Heritage Library, BHL），WoRMS（全球海洋物種登錄系統）、FishBase（魚庫）、OBIS（海洋生物地理資訊系統）、AquaMap（水生生物圖庫）等等。如何規劃、整合及推展這門學科，其重點及優先工作項目，在 2009 年 6 月初於倫敦所舉行的一項 e-Biosphere Conference 中已有若干決議及共識。

二、 兩岸生物多樣性資料庫之建置與進展

台灣自 2001 年起在國科會及農委會的支助下，由中研院生物多樣性研究中心負責召集，目前已順利完成物種名錄（TaiBNET）（http://taibnet.sinica.edu.tw/）之建置。另在數位典藏國家型計畫（TELDAP）之贊助下，目前已整合了台灣絕大多數標本館所典藏數十萬筆的標本及其相關資訊的數位化資料。上述這些資料庫之交換標準 Metadata 及原始出現資料之欄位，乃至分類系統均係隨著國際上之 COL 或 GBIF 所採用之國際標準及相關整合協定如 Darwin Core、DiGIR 或 TAPIR、IPT 等等，逐步推動，目前均尚稱順利。

中國大陸在過去十餘年來所努力建置及整合之資料庫或資訊網亦甚多，如中國生物物種名錄（http://www.sp2000.cn/），國家標本資源共享平台，中國數字植物標本館，中國自然標本館（http://www.cfh.ac.cn），及種質資源庫等等。目前大陸負責此項工作的單位較多，動物資料由中科院動物所負責，植物資料由中科院植物所負責，微生物資料由中科院微生物所負責，生命條形碼資料庫則可能由昆明動物所、植物所及中科院微生物所等分工合作；另還有保護區、種質資源及長期生態或大樣區之監測資料等等，各有不同的單位負責。而這些生物多樣性相關資料之統整，在目前主要應由大陸地區的 DIVERSITAS 的中國委員會來負責。其主任委員爲中國自然科學基金會主任陳宜諭擔任，植物所馬克平所長爲執行秘書。中國大陸迄今雖未參加 GBIF 之組織，但在最近這三年都分別與 FishBase, Species 2000, BOL 及 EOL 正式簽約合作，投入大量經費、人力、物力來推動大陸生物多樣性資訊之整合及與國際接軌。包括鏡像站的建立、資訊交換平台的建立，資料之收集與整合，以及教育、訓練及推廣等工作。

三、 兩岸生物多樣性資料庫合作之倡議

由於海峽兩岸之生物多樣性或生物相（區系）在起源、形成、種化、擴散、分布及維繫上關係最爲密切，在研究、保育、教育及資源的管理與永續利用上更需要兩岸交流合作，才能有更好的成果，而資料庫之合作更是其他各項研究項目的基礎建設工作。也因此生物多樣性乃被納入 2010 年起應開展之兩岸合作的科研究項目。爲研商生物多樣性未來合作之重點工作項目，2009 年 5 月在昆明舉行了一項「海峽兩岸生物多樣性學術交流研討會」，會後共獲得七項合作項目之共識。其中的生物多樣性資訊及生命條碼的整合項目已列爲未來主要的合作項目之一。由於海峽兩岸在資料庫之建構上早已各自累積相當豐碩的資料，也各自建構完成完整的資訊架構及網頁，因此資料庫之整合與分享應不困難，是最容易取得及展現合作成果之項目。有鑑於此，台灣代表乃於 2009 年出席在 8 月 27～28 日兩天在內蒙古錫林浩特由馬克平所長所主辦的「生物多樣性資訊學會議」，與來自大陸各地從事生物多樣性資訊的工作者研商，會後並至北京與負責植物資料

的覃海寧、負責動物資料的喬慧捷做深入的交流。10 月 13 日也趁著赴南非開普敦參加 DIVERSITAS-OSC2 會議，與對岸負責動物資料庫的紀力強研究員研商下一步合作內容。11 月 9 日也趁著赴墨西哥市參加生命條碼年度會議，與對岸負責微生物及條形碼資料庫之馬俊才研究員等人之協商微生物資料合作事項，初步就未來三年擬優先進行的工作、進行討論並取得共識後，本已擬妥計畫書準備申請兩岸合作之專題計畫，但因技術性問題而未能即時提送。

四、 擬優先進行之合作項目

（一）「兩岸物種名錄」之整合及條列

生物物種學名部份，兩岸可就中文名（台灣已有五萬種，中國已匯集六萬多種），做出對照表並列出來。共建「中華物種名錄資料庫」或稱「海峽兩岸物種名錄資料庫」網站。進而以此模式再推廣到生命大百科、生物誌、生物多樣性遺產圖書館（BHL）之資料庫之整合等等。但此工作首先必須先就兩岸所採用之資料庫標準、必填欄位等先做洽商及整合。當兩岸採用同一標準後，各項資料庫之整合與分享自然水到渠成。

「中國物種名錄」於 2008 年經由與 Species 2000 合作，建構 Catalogue of Life, China（http://csvh.org/cnnode），目前收錄動物 8,804 種、植物 40,744 種，合計 44,140 筆有效學名，外加 74,658 筆同種異名，此名錄由中科院動物所負責維護。

「台灣物種名錄」於 2001～2003 年在國科會的支助以及超過 100 位台灣分類學家的通力合作下完成，並建立 TaiBNET 網站（http://taibnet.sinica.edu.tw）提供完整的網路查詢功能。2004～2006 年雖然沒有計畫經費支助，中央研究院生物多樣性研究中心仍然繼續進行名錄的增修工作，並於 2005 年出版 Annual Checklist 光碟，2007 年月起，國科會再度恢復支助三年計劃。迄今已收錄，病毒 495 種、細菌 435 種、古菌 6 種、原生生物 1,377 種、原藻 1,953 種、真菌 5,817 種、植物 7,060 種、動物 34,149 種， 總計：1+7 界 58 門 133 綱 661 目 2,959 科 16,763 屬 51,519 有效種，外加五萬筆以上之同種異名，大致上已經涵蓋台灣

地區大部分的物種，目前仍持續增修。

（二）「兩岸生物標本資料庫」及條碼之整合

整合對象包括一般標本、酒精保存的組織標本、冷凍遺傳物質標本及其存證標本（voucher specimen）在內。由於條形碼之資料整合，目前國際上已有 GenBank 及 BOLD（Barcode of Life Database）等大型全球基因資料庫，並有完善之分析及比對的軟體工具，所有條形碼之資料如入庫並上網公開則全球使用者均可查詢及使用，因此兩岸如另建一個兩岸的條形碼資料庫，只能供兩岸科學家使用，在技術上、應用性、功能性、及必要性上均值得商榷。然而，條碼庫要建立完整，最困難的瓶頸在於，難以採到所有物種的新鮮標本，而全球各地有將近 1/3～1/2 的物種均爲稀有種或罕見種，甚至瀕危或滅絕，使得及早採集及典藏組織及其存證標本要比定序獲得條碼更爲重要。有了組織標本隨時均可定序得到條碼，且不只可作動物之 CO1，植物之 rbcL 及 mtbK，或真菌的 ITIS 等基因而已，任何基因片段均可獲得。也因此酒精保存的組織標或冷凍遺傳標本及其存證標本之典藏資料之合作應該比只有 CO1 等條碼的合作更重要更有用。此外，標本資訊的公開較不涉智財權或使用優先權的問題，借閱或索取標本需經採集者、提供者或管理者之同意才可取得，經由磋商更能增加雙方合作之機會。反之條形碼資料一旦上網公開即無法控管，除非只作宣告有該物種之條碼而暫不公開。但此舉實與提供組織標本之資訊相當，因只要有組織標本，隨時可作條碼，加上定序之技術及儀器愈來愈進步，成本愈來愈低，如果沒有立即的需求，實在不該急著定序。

大陸標本數位化工作主要以植物標本爲主，目前中國數字植物標本館（http://www.cvh.org.cn）已經收錄超過 2,850,000 筆標本資料，1,500,000 張標本照片，中國數字植物標本館網站由中科院植物所負責維護。

台灣標本數位化工作主要以數位典藏計畫爲主，除了可以由各個數位典藏計畫網站及數位典藏聯合目錄查詢外，目前已經將大部分的標本資料整合至 TaiBIF 網站（http://.taibif.org.tw）並與 GBIF 接軌。總計約 400,000 筆資料，分屬於 14 個單位（中研院生物多樣性博物館、行政院農委會林業試驗所植物標本館、行政院農委會林業試驗所昆蟲標本館、國立臺灣大學植物標本館、國立臺灣大學魚類

標本館、國立臺灣大學昆蟲標本館、國立自然科學博物館、行政院農委會特有生物研究保育中心、行政院農委會水產試驗所、國立海洋生物博物館、國立海洋科技博物館、國立清華大學、國立臺灣博物館、國立台灣海洋大學、生物資源保存及研究中心），主要以植物標本、魚類標本及昆蟲標本爲主，TaiBIF 網站由中研院生物多樣性研究中心負責維護。至於生命條碼及冷凍遺傳組織標本之典藏在農委會之支助下，亦已逐年累積並建立專屬網站（http://cryobank.sinica.edu.tw）。目前已累積近 4,000 種，近萬件之組織標本，並對外公開。

（三）「兩岸生物多樣性學科技名詞對照」上網

建立可以同時顯示台灣及大陸譯名的網頁，以及建立協調新名詞中文翻譯的討論平台。由於各項與生物多樣性相關之文件，或是科普文章均需中譯，如 Gbit、EOLETTER、FishBase2000 及 GBIF 之若干文件等。如果哪一方已先進行翻譯，則另一方則不需要再重覆進行翻譯，以節省人力物力。如 GBIF 之通訊 GBit、FishBase2000 已由台灣中譯，只須將台灣已譯妥之正體字版直接改爲簡體字即可，同樣 EOLETTER 通訊大陸已有簡體字版，我們直接改爲正體字，並說明出處及來源即可。但其中不少專有名詞兩岸之譯法多有所不同，如 information（資訊 v.s 信息）、EOL（網絡生命大百科 v.s 生命大百科）、GBIF（全球生物多樣性信息網絡 v.s 全球生物多樣性資訊機構）等等，需要建立詞典庫以便電腦可以自動將這些不同中譯的名詞作轉換。因此先建立此一詞典庫十分重要。此項工作也正好符合兩岸刻正推動的《中華大辭典》或《海峽兩岸學術名詞對照》工作之需求，主動提供生物多樣性學科海峽兩岸上千條的詞條。

五、 進行之步驟與方法

（一）合作建構兩岸物種名錄

1. 合作基礎及努力方向

由於大陸與台灣都與 Species 2000 密切合作，並盡可能採用 Species 2000 的資料結構，使得兩方的名錄資料結構極爲相似；目前兩方也都不約而同的使用

PHP+MySQL+JacaScript 架構網站及資料庫，使得未來的資料整合工作，展現非常有利的基礎，唯未來的資料整合工作仍有幾項挑戰有待克服：

（1）一些分類群，兩岸採用不同的分類系統，無法將資料整合至單一分類樹。

（2）一些有效學名的認定不同，部分有效學名被對方歸類爲同種異名，也有部分同種異名被對方歸類爲有效學名。

（3）大陸物種名錄資料結構與 Species 2000 完全相同，台灣物種名錄的資料結構，是參考 Species 2000，並針對我們自己的需求設計，與 Species 2000 並不完全一致，特別是有關中文名的部分，台灣物種名錄的每一個分類階層都有相對應的中文名欄位，每一個學名都有標準中文名及俗名欄位，而 Species 2000 則單純地一律以「地方名」待之。

（4）台灣物種名錄爲配合 Ajax 非同步資料傳輸技術，專門設計的資料結構，可否與大陸整合，有待進一步探討。

（5）兩方使用完全不同的資料維護平台（後台軟體），有待協調出一致的資料維護協定。

2. 合作內容及工作項目

（1）雙方各派 1 名分類專長、2 名資訊專長的專家合組工作團隊，共同負責建置資訊平台（前端網頁及後端維護系統），於北京及台北，各建置一個網站及資料庫，網站將提供中文正體、中文簡體及英文三種語言，北京及台北的網站內容完全相同，資料庫彼此同步。

（2）對於無法取得一致結論的分類群，預訂設計不同分類系統之間的轉換機制，提供給使用者選擇使用，此部分資訊技術可透過 Ajax 非同步資料傳輸技術解決。

（3）於北京及台北各召開一次整合會議，邀請兩岸分類學者共同參與討論，協調一致的作法。

（二）合作建構兩岸數位標本館

1. 合作基礎及努力方向

標本資料具有相當多的共通性，不外乎採集資訊、鑑定資訊及保存資訊，儘管各單位建置的標本資訊結構各異，仍可透過將資料匯出至標準格式，達到整合的目的，也可以利用現成的資料交換軟體（例如 DIGIR、TAPIR）及資料交換標準（如 Darwin Core）達到分散整合的需求，唯未來的資料整合工作仍有幾項挑戰有待克服：

（1）台灣已經普遍接受 Darwin Core，作爲標本資料的共通標準，但 Darwin Core 尚未能獲大陸各標本館接受。

（2）Darwin Core 並未考慮語言問題，應適度修改，把中文欄位考慮進去。

（3）許多標本沒有採集點座標，由地名轉換成座標的工作將是一大工程。

（4）標本採集點座標單位紛亂，統一座標單位將是當務之急，轉換工作亦將是一大工程。

（5）同一物種的標本使用的學名可能不同，必須等到兩岸物種名錄整合完成後，才能進一步依照名錄，找到相對應的標本。

（6）資料量龐大，需有充足的網路頻寬。

（7）許多尚未鑑定到種標本，資料難以整合。

2. 合作內容及工作項目

（1）雙方各派 2 名資訊專長的專家合組工作團隊，共同負責建置資訊平台（前端網頁及後端維護系統），於北京及台北，各建置一個網站及資料庫，網站將提供中文正體、中文簡體及英文三種語言，北京及台北的網站內容完全相同，資料庫彼此同步。

（2）待兩岸物種名錄整合完成後，再將標本資料進一步與名錄整合。

（3）於北京及台北各召開一次整合會議，邀請兩岸各標本館代表共同參與討論，協調一致的作法。

六、 預期成果

第一年：

1. 完成兩岸具合體作之架構、方法、優先項目、期程細節之協商，乃至協議書或備忘錄之文字內容之研擬。
2. 藉互訪及舉辦工作小組會議或工作坊，完成兩岸物種名錄整合所需之資訊技術及交換標準之交流及統一。
3. 完成兩岸物種名錄資料之整合。
4. 完成「中華物種名錄資料庫」或「海峽兩岸物種名錄資料庫」之網頁製作，並對外公開。
5. 蒐集兩岸生物多樣性學科領域之科技名詞之中譯名，進行其譯名妥適性之專家討論，建立詞庫（字典）。

第二年：

1. 繼續增修訂兩岸物種名錄及科技名詞。
2. 藉互訪及舉辦工作小組會議或工作坊，完成兩岸生物標本含組織標本整合所需之資訊技術及交換標準之交流及統一。
3. 完成兩岸生物標本資料之整合。
4. 完成「中華生物標本資料庫」或「海峽兩岸生物標本資料庫」或虛擬博物館之網頁製作，並對外公開。

第三年：

1. 繼續增修訂兩岸物種名錄、科技名詞及標本資訊。
2. 挑選兩岸生物多樣性研究及科普相關之重要文件或資訊納入資料庫（正、簡體字版）。
3. 兩岸物種解說網頁，含影音資訊之整合與資料交換標準之訂定。
4. 擴大舉行研討會，檢視三年合作成果並規劃未來合作項目。
5. 完成「中華生物多樣性資料庫」或「海峽兩岸生物多樣性資料庫」或之網站製作，並對外公開。

【作者簡介】

邵廣昭　男

職　　稱：中研院 生物多樣性研究中心 研究員

職　　務：中研院 生物多樣性研究中心「系統分類及生物多樣性資訊專題中心」執行長

研究領域：魚類分類、海洋生物、生物統計、生態與進化

個人簡介：美國紐約州大 Stony Brook 生態進化系博士。回台後即服務於中研院動物所，1991～1994 年出任首任海洋大學海生所所長，1996～2002 年接任中研院動物所所長。之後，負責推動中研院生物多樣性研究中心之設立，並於 2004～2007 年代理中研院生物多樣性中心主任。曾任中國生物學會理事長及自然生態攝影學會理事長，台灣魚類學會理事長、鯨豚協會理事長，並兼 DIVERSITAS 及 GBIF 等委員會台灣委員會之執行祕書。目前擔任行政院永續會委員及農委會生物多樣性推動小組委員。長期推動台灣魚類及海洋生物之研究、教育及保育工作不遺餘力，目前正積極推動台灣生物多樣性資訊之整合工作。除發表百餘篇學術期刊論文外，專書亦甚多，曾獲金鼎獎、十大傑出青年、生物科學研究獎、國科會傑出研究獎、教育文化獎章、侯金堆學術研究獎、十大傑出農業專家等。代表性著作為《海洋生態學》、《台灣常見魚介貝類圖說》、《魚類入門》、《魚類圖鑑》、《藍色運動》等書。

聯絡電話：886-2-27899556 ext.200；886-2-27899545 ext.200

聯絡郵箱：zoskt@gate.sinica.edu.tw