

# 太魯閣國家公園研究資訊目錄系統之建立

金恆鏞<sup>1</sup>，鄭美如<sup>2</sup>，林朝欽<sup>3,4</sup>

<sup>1</sup>亞熱帶生態學學會；<sup>2</sup>林業試驗所技術服務組；<sup>3</sup>林業試驗所森林保護組；

<sup>4</sup>通訊作者 E-mail: [chin@tfri.gov.tw](mailto:chin@tfri.gov.tw)

**[摘要]** 為建立太魯閣國家公園研究資訊目錄及研究資料管理之完善，不再因委託計畫結束後導致資料流失的情形發生，並預期實現研究所蒐集的原始數據可再利用、延續數據生命與提昇數據價值之目標。本研究利用生態資訊學所發展的資訊管理系統架構，制訂共通之數據格式以整合異質性資料來源，期能涵蓋動物、植物、地質、人文等不同類別，完成建置太魯閣研究資訊目錄，透過元數據的建置，將原始資料及其文件長期儲存於資訊管理系統中。本研究計已完成蒐集 1983-2008 年間 203 份太魯閣國家公園過去委託研究計畫結案報告書，並依據生態元數據語言 EML，將過去委託研究計畫之一般資訊建置完成符合國際資訊交換標準的文件資料集。此外，並辦理兩場說明與教育訓練研習會，協助 18 個計畫進行 EML 文件編輯，未來將有助於國家公園管理處訂定保育研究之發展及應用於資源經營管理工作。本研究並提出建立太魯閣國家公園研究計畫繳交資料之元數據規範及標準流程之建議，作為未來研究資料管理與長期倉儲之參考。

**關鍵字：**元數據、資料管理、資訊、長期生態研究

## Building a metadata catalog system for Taroko National Park

Hen-Biau King<sup>1</sup>, Meei-Ru Jeng<sup>2</sup> and Chau-Chin Lin<sup>3,4</sup>

<sup>1</sup>Society of Subtropical Ecology, <sup>2</sup>Division of Technical Service, Taiwan Forestry Research Institute, <sup>3</sup>Division of Forest Protection, Taiwan Forestry Research Institute, <sup>4</sup>Corresponding author E-mail: [chin@tfri.gov.tw](mailto:chin@tfri.gov.tw)

**ABSTRACT** A well-built research data catalog system can facilitate researchers to store their raw data effectively and to share the data among their colleagues readily. The system makes possible of reusing archived data by avoiding data entropy in the future. This research provides service to researchers to manage their data easily and effectively. This system was particularly designed for the Taroko National Park. The system allows storage of metadata and raw data collected by individual researchers or agencies from the Park. The data catalog system adapts the ecoinformatics framework and Ecological Metadata Language (EML) standard. The created metadata from research's raw data were stored in a metadata database based on the EML standard. A website has been created with query and browse functions for potential users to locate data. Data from 203 projects over 25 years (from 1983 to 2008) were archived. Two EML training workshops were held to promote and help scientists use this data catalog system. Recommendations for future data management were suggested.

**Keywords:** metadata, data management, information, long term ecological research

## 前言

國內各國家公園從創設以來，為了進行規劃、保育或經營等目的而進行許多委託或自行研究的計畫，涵蓋範圍從地質水文、植物生態、動物生態、人文史蹟等等不一而足，多年下來已累積了為數不少的研究計畫，目前雖然能從國家公園的網站上搜尋到這些計畫的成果報告書，但均屬於成果報告文獻資料，文獻所依據的原始數據(raw data)卻未儲存於委託單位，而這些數據往往隨著研究計畫結束後逐漸流失，即使資料留存在原先的研究人員手中，仍舊無法累積成為國家公園內的基本資料，更難以提供進一步的研究與資訊分析而供經營管理者或決策者使用。由於許多資料因缺乏有系統的管理而流失，不但減損了這些資料的價值，同時也相對提高了研究成本，更無法整合既有資料，產生新的研究價值。因此如何在龐雜、分散與異質性的資料中確保可以長久保存與取得，以及如何發揮資料的分享與整合功能，必須要有一個共同的描述標準。更甚者，這樣一個標準需能協助研究人員妥善管理自己的資料，從資料開始搜集前的設計到進入倉儲後的分享機制均需考慮，才能讓資料不再因計畫結束後即流失(金恆鏞 2006)。此為了達成前述資料管理的目的，描述資料內容的標準必須被所有研究者遵循，才能進行資料的存取、索引、檢索與共享，metadata (中文翻譯有詮釋資料、後設資料、中繼資料、元數據等，本文採用元數據)便是扮演這樣的角色與功能，元數據導向即是在一個標準化格式下描述所蒐集資料內容的標準。因此，元數據標準的選取、推動與實施是研究資料管理不可或缺的一環(林朝欽等 2006)。

生態資訊學(ecoinformatics)為整合生物資訊學(bioinformatics)與生物多樣性資訊學(biodiversity informatics)兩個資訊學門所發展的新領域(Cushing et al. 2003; Michener

2006)，以處理複雜的生態資料為主要考量，所發展的資訊管理系統與工具，亦朝向整合生物資訊學與生物多樣性資訊學，使此生態資訊學成為可分享資料的大系統。美國國家生態分析與整合中心(National Center for Ecological Analysis and Synthesis, NCEAS)與國際長期生態研究網的生態資訊管理委員會(International Long Term Ecological Research IM Committee)建構的生態元資料語言(Ecological Metadata Language, EML)與科學工作流程系統 Kepler 可以說是此學科的代表，且已初步完成將基因整合到生物圈的資訊系統架構(林朝欽等 2008; Michener 2006)。

為了改善這種資料流失的情形，太魯閣國家公園管理處於 2009 年委託亞熱帶生態學學會建構該處委託研究之資訊目錄系統，除 2009 年度該處委託之相關研究案需遵循元數據標準建置研究計畫內容描述與原始數據儲存外，並藉由回溯過去研究報告書之蒐集，期能建立完整的研究資訊目錄，更希望未來將這些研究之原始數據長久保存，透過詮釋元數據將相關資料與其他跨領域或國際型研究接軌。建立這種符合國際規格的研究資訊目錄，將有助於日後對資料的存取與利用，並促進資料的分享與整合，以提供國家公園及保護區的決策者所需的資訊。

本研究預期解決目前國家公園內資料分散的困境，除協助委託計畫研究人員可以妥善管理資料，建立元數據的分享機制，不再因委託計畫結束後導致資料流失的情形發生外，並實現資料再利用、延續資料生命與提昇資料價值之目標。讓研究之原始數據能繳交至委託單位，除可協助各委託計劃做為資料備份(異地備份)，也將有助於國家公園管理處未來保育研究之發展及應用於資源經營管理工作，為達成上述的目標，本研究採用生態資訊學理論與技術建構此一系統。

## 材料與方法

### 一、資料描述標準及介面工具之設計

元數據主要是資料內容的描述，因此可以定義為資料的資料，元數據導向即是一個標準化的格式下，描述所蒐集到的原始數據與相關資料內容的標準。因此本研究依生物多樣性與生態研究之特性，對現有的元數據標準進行分析及設計出可結合不同標準的架構。並據以蒐集可用的介面工具加以修改，以提供未來系統使用。

### 二、資訊目錄系統建置

本研究以所採用的元數據標準為依據建構系統，系統架構分為軟體介面、認證與倉儲三部分。軟體介面以可跨作業平台(Windows、Linux、Mac)的自由軟體或語言為主，再導入適用於中文系統的編碼，認證機制亦以開放原始碼的系統及工具，並提供使用者自訂密碼機制，後端倉儲則選擇可存放元數據及原始數據的資料庫管理系統。此架構可彈性成為集中式的中央管理系統，亦可為分散的聯合管理系統，若採前者則只有單一集中系統管理單位，若採後者則可有許多管理單位，但透過合作協議聯合成為一個

網狀系統，各自獨立運作但可互相取得(harvest)各系統內容。

### 三、系統操作教育訓練

為協助研究人員了解系統及以建置元數據，本研究計舉辦 2 場次之教育訓練，讓研究人員可以順利建置其計畫之元數據，存放於本機或測試系統中，未來系統完成後便可以順利上傳至太管處資訊管理系統中做保存、查詢與分享。

### 四、元數據建檔

蒐集與彙整太魯閣國家公園 1983-2008 年來經營管理相關計畫與論文，完成太管處歷年來委託研究案相關元數據建檔，並聯繫相關計畫主持人能否提供原結案計畫之原始數據檔。

## 結果與討論

本計畫完成蒐集 1983-2008 年度間 203 份委託研究之成果報告書，並依動物、植物、地科、人文等項目分類，各類資料份數如圖 1 所示，其中以動物、植物、地科等類佔了委託計畫案的大宗。建置完成太魯閣國

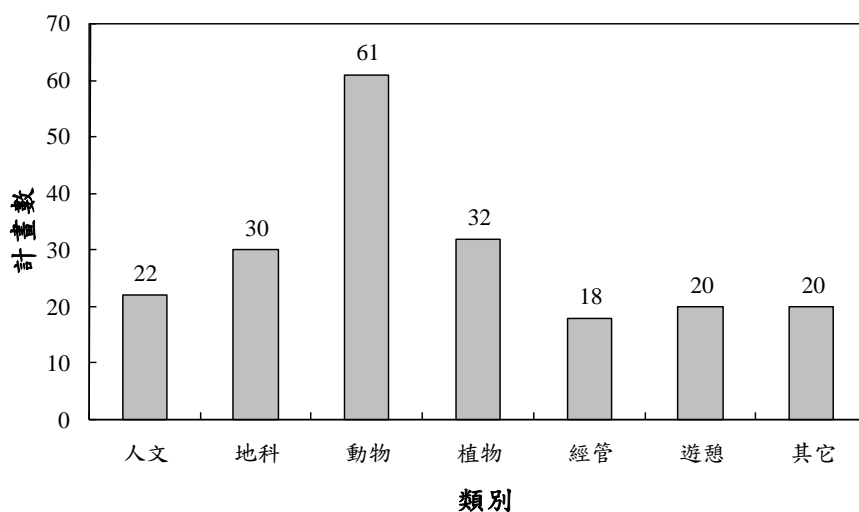


圖 1. 203 份歷史計畫報告書類別份數



圖 2. 建置太魯閣國家公園管理處研究資料目錄



圖 3. 輸入完成 203 份的歷史資料集

家公園管理處研究資訊目錄(圖 2)，網址 <http://metacat.taroko.gov.tw/taroko/>，除可利用查詢介面搜尋相關資料外，系統並提供資料文件檢覈工具，將可確保資料品質之良窳。本計畫業已輸入完成全數 203 份的歷史資料集(圖 3)，內容除了計畫摘要、關鍵字詞外，也將大部份計畫的研究方法相關資訊輸入建檔，部分物種資訊亦詳細說明，大幅提高資料的利用性。

此外，已於 2009 年 7 月 15 日及 11 月 9 日辦理二場次 EML 教育訓練，2009 年度太魯閣國家公園管理處委託的 18 個計畫皆有派員參加，人數逾 50 人，系統目前已可檢索到 2009 年度該處委託的相關研究案。訓練課程內容涵跨從資訊管理的基本概念，到進階的資料管理系統操作，以及未來資訊技術發展的最新趨勢。並完成編撰元數據編輯系統操作手冊一本，已於教育訓練課

程中分送，並檢附光碟版供參，可供研究人員自行參考與按照步驟操作，大幅提昇研究人員日後自行建置的能力。

本研究經比較各元數據標準後，選擇 EML 作為系統內所有生態資料的標準，除因 EML 具有下列所述優點外，更因以 EML 為標準所設計的編輯工具亦已開發出多年，這種使用的軟體介面是其他標準所欠缺的(陸聲山、林朝欽 2005)。EML 由美國生態協會(Ecological Society of America)和相關附屬協會的努力而奠下基礎，從最早 1997 年的 EML 1.0.0 版問世後，中間經過多次的版本修正，目前為 2.0.1 版，此標準具有以下的 4 個特點：

1. 模組化(modularity)：EML 以不同模組的集合來代替單一龐大的標準架構，因此可以因應廣度和深度上的成長。模組間具階層關係讓數據或相關資料的描述可由簡單到詳盡，EML 共提供 22 個模組來描述一份數據或相關資料，若有特殊領域超出這 22 個模組範圍，使用者可以自己再定義模組加入。
2. 詳細的結構(detailed structure)：EML 的設計以高度結構化的元素(element)，記錄使用者所輸入的數據或相關資料的描述，這種結構是可被機器辨認與人類閱讀懂得。
3. 相容性(compatibility)：EML 對數據或相關資料的描述的句法，依循可延伸標記語言(eXtensible Markup Lanaguage, XML)的技術，大幅增加 EML 文件與其它元數據語言之間的轉換與相容。
4. 強制型態(strong typing)：EML 的文件是以 XML 為架構，因此 EML 的句法符合 XML 語言的定義及規則。XML 架構是一個符合網際網路的共通標準，因此符合 EML 句法的元數據文件結構，除可以在網際網路上進行交換，也提供資料檢驗的能力。例如，若輸入變數為日期(date)的資料時，這些輸入的數值將根據 XML 架構中的日期定義再次被檢驗。

由於 EML 可以讓使用者定義擴充模

組，因此為標本或物種數據描述的 DwC 即可鑲入 EML 的文件中。DwC 由物種分析專案(The Species Analyst Project)研發，提供 48 項元素作為一份記錄標本的所有訊息(Canhos et al. 2004)。本研究將此標準以數據表單的方式加入 EML 的結構中，成為可以結生物多樣性與生態數據或相關資料的描述的共通標準。

元數據的編輯工具選擇以 EML 標準所設計的 Morpho，此軟體是專為生態學家設計的一個數據或相關資料的描述管理的共享軟體，提供了一個容易使用與跨作業平台應用的介面，使用者得以在本機和網路上取得和操作元數據及原始數據(林業試驗所 2008)。

Morpho 使得生態學家可以輕易的產生符合 EML 標準的元數據，並且也建立元數據目錄，能讓使用者查詢(query)、編輯(edit)和觀看(view)已有的數據或相關資料的描述。Morpho 基本的操作功能包含有：產生和編輯元數據、搜尋和查詢元數據彙集、觀看資料與元數據彙集、校對和編輯資料與元數據、存取控制等。Morpho 所儲存的資料類型，包括了像數據或相關資料描述的表單和使用者自行定義的內容。資料套件(data package)是 Morpho 產生資料檔案及其描述的文件(document)的精靈，可幫助研究者產生最低標準的文件，這個過程包括 15 個簡單的步驟，每個步驟都利用不同的視窗給使用者輸入。套件精靈在必須要填寫的欄位利用紅色標識出來，每一個步驟也有填寫的說明與建議，加上還有線上的使用手冊提供，使它成為相當簡易好用的介面。

Morpho 為美國 Knowledge Network for Biocomplexity(KNB)計畫為長期生態研究網設計的 EML 文件編輯軟體，但因其設計之初未考慮多語系國家之使用者，且本計畫需整合中文相關之資料，因此將之修改為中文文化使用介面如圖 4，以利未來國內建置與保存資料之用。

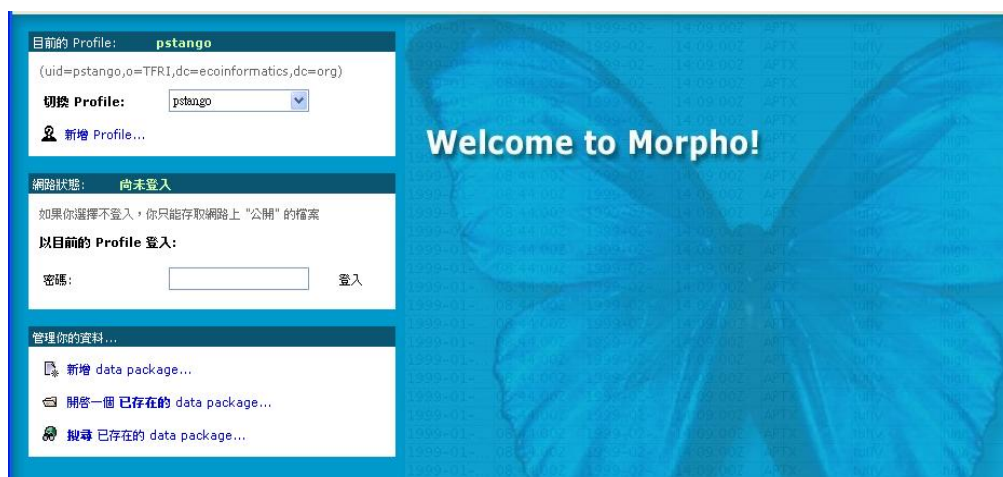


圖 4. 中文化之 EML 資料編輯介面

EML 文件的資料編輯完成後研究人員可以儲存在自己的電腦中，並藉由 Morpho 管理，亦可透過網路系統存入 EML 文件資料庫，此資料庫系統並命名為元數據目錄 (Metadata Catalog，簡稱 Metacat)。Metacat 是利用 Java 程式語言建構而成的，它所使用的資料庫為一般常見的關聯式資料庫，使用者需經身份認證存取資料 (Lin et al. 2006)。在資料展示方面，Metacat 利用網頁的文件型式，將資料呈現給使用者 (圖 5)。除以網頁展示外，也可以使用 Morpho 來讀取 Metacat 的資料 (圖 6)，一份完整的 EML 文件，包含多筆原始資料檔案。

Metacat 的主要功能是作為 EML 文件的倉儲，其中 EML 文件被儲存於關連式資料庫，目前可支援的資料庫為 Oracle、PostgreSQL 以及 MS SQL，本研究採用共享軟體的 PostgreSQL 作為後端倉儲的資料庫管理系統，而原始數據集如物種名錄則是被儲存於檔案系統中。Metacat 與 Morpho 具有主從式的 (client-server) 關係，Metacat 是位於提供服務端的伺服器，而 Morpho 是處於提出服務要求端的軟體。Morpho 可以對 Metacat 提出的服務要求包括了進行上傳、下載、刪除以及查詢等四種服務要求，此種架構提供了個人與社群間可分享研究資料

的良好機制，不論研究者目前是否願意分享資料，都可以有一個良好的工具來管理研究資料。

資訊管理系統的建置是資訊分享的第一步，也是資料儲存與分享最重要的骨架，而資訊管理員便扮演著建構與維護的靈魂角色。由於資料層面的龐雜，一般而言，生態資訊管理系統又較工商業的資訊管理系統複雜，且受限於市場需求小，因此建置成功的生態資訊管理系統與維護是必須透過生態資訊管理員、資訊技術人員、田野與實驗室技術員、研究人員，在考慮研究經營管理目標後相互協調討論而成。生態資訊管理系統建置的目的，是用來支持研究網絡與科學性社群間的互動，並使科學性社群與公眾可以容易的存取資料與元數據來相互分享，同時確保後代的整合與使用，因此應有自己獨立的經營管理計畫，並藉由內部與外部的檢視，以評估管理系統的目標是否達成。本研究應用生態資訊學發展的最新技術，以達成資料共享、資訊整合，將可促進未來自動化資料分析與整合，以達成知識管理之最終目的。

EML 是一種根據生態學科而發展出的元數據規格，國內由林業試驗所率先採用並推廣至亞太地區的長期生態研究網。





圖 5. 完整的元數據與資料集



圖 6. 資料欄位屬性詳細填寫說明

這套標準可作為生態資料庫連結與整合的共同標準，目前國際長期生態研究網 (ILTER)與 GBIF 的代表已共同討論這兩大機構可能的合作機制，嘗試將長期生態研究與生物多樣性資料庫的服務接軌，以加速長期生態研究與生物多樣性的資訊交流與資料分享。林業試驗所在生態資訊管理這方面的努力，結合 EML 與 Ecogird 的分享系統，

完成自動化的工作流程 Kepler，除可以應用在生物多樣性指數運算與物種分佈模式預測，大幅提升研究效率(蕭其文 2006)。此外，透過分散式儲存機制，除能將大量的資料做更有效率的運用與備份，未來仍將繼續研究，朝向透過 EML 將資料做進一步的加值運用。

可預見的未來，EML 這一標準已打破國

界的限制，將成為新一代資料整合與分享的標準。對研究人員來說，到野外辛苦的蒐集資料並不是一件困難的事，但所蒐集到的資料如何能在日後發揮更大的效益，而不是當研究完成後，發表一兩篇報告後資料的價值便降低甚至消失。EML 的規劃與實施，將是生態資料分享與整合最基礎的資訊建設工作之一，可使得以往生態研究之資料處理、分享與整合的困難逐漸獲得改善，也打破以往個人蒐集資料、儲存及分析使用後資料流失的困境，使得資料之再利用及創造更珍貴的價值得以實現。

有效的資料庫整合方式是在研究過程的一開始，各個研究人員當提供他們的資料做為資料倉儲或歸檔用時便能符合 EML 的規格與相關技術。很不幸地，研究人員對資料庫的用途與使用卻存在許多的障礙有待克服。此外，這樣的整合技術將可能造成研究人員在研究方式上產生許多改變，而造成實行上的困難或配合意願低落。因此要發展出適當的技術，將必須對研究過程有詳細的了解，同時也需要資訊人員、資訊管理員和研究人員或資源管理人員等三方之間密切的合作才得以完成。

雖然資料長期的價值將取決於我們完善的元數據記錄與管理方式，但這種符合 EML 規格的資料歸檔工作，卻是非常耗費時間與精力的工作，也將是我們面臨資訊整合的最大瓶頸(Cushing et al. 2003)。雖然生態資料的複雜性、異質性使得資料整合困難重重，但面臨資料分享與整合所產生的好處與效益，將促使研究人員樂於這種轉變才對。

中央研究院邵廣昭等(2008)多次建議政府各委辦單位應該要在委託計畫簽合約時，規定計畫結束後需要繳交原始資料，以利數位化資料保存。太管處為政府部門率先實施原始資料管理的單位，後續的維持仍有待持續努力。國科會目前亦正推動此制度與研議機制，預計明年度起國科會委託之計畫，將簽約規定需要繳交原始資料。太管處

嘗試建立委託相關研究案需遵循元數據標準的規範，任何制度施行之初，恐將或多或少遭遇阻力，但從建構長期生態元數據庫目錄做起，採循序漸進方式，若能行之有年，必將看出其績效。

本研究雖將歷年來委託研究案相關元數據建檔，但部份委託計畫案已因年代久遠，計畫主持人已退休或離職，恐將增加連繫上之困難，透過網路搜尋已更新部份計畫主持人之現職資訊，將可促進未來作為聯繫相關計畫之原主持人，洽詢能否提供原結案計畫之原始資料檔之需。

## 結論

未來研究資料管理與長期倉儲之建議，宜區分為新提計畫與舊有歷年研究計畫兩部份。新提計畫可依 1. 基本可辨識；2. 資料可被發現；3. 資料可被搜尋；4. 資料可被擷取；5. 資料可被整合分級建立，再依合約規定項目分別對應，具有彈性，可供太管處具體要求計畫所需完成之具體內容。而詮釋資料舊有歷年計畫，可參採國科會數位典藏計畫編列盤點經費，而逐步尋回舊有計畫之原始資料。

## 誌謝

本研究進行期間承太魯閣國家公園保育課同仁與東華大學自然資源研究所師生之諸多協助；2009 年太魯閣國家公園委託研究案之計畫主持人及其團隊參加兩次研習營之熱忱；林業試驗所生態資訊團隊之技術協助，謹此一併致謝。

## 引用文獻

林朝欽、鄭美如、陸聲山。2006。生態資訊學研究與生態資訊管理之推展。林業研究專訊 13(6):4-8。



- 林朝欽、鄭美如、陸聲山。2008。生態資訊學之發展與應用。台灣林業科學 23(S):1-10。
- 林業試驗所。2008。Morpho-生態後設資料語言編輯系統。林業叢刊第 184 號。
- 邵廣昭、賴昆祺、林永昌、柯智仁、陳麗西、李翰、林欣樺。2008。數位典藏計畫中生物多樣性資料之整合。昆蟲與螞蟥標本資源之管理與應用研討會專刊：233-244。
- 金恆鏞。2006。建置生態資訊管理系統-兼論開放原始碼與黃金企鵝獎。林業研究專訊 13(6):1-3。
- 陸聲山、林朝欽。2005。生態資料檢索標準與使用— EML (Ecological Metadata Language)介紹。林業研究專訊 12(1):5-8。
- 蕭其文。2006。生態後設資料之進階應用系統。林業研究專訊 13(6):20-23。
- Canhos VP, S Souza, R. Giovanni and DAL Canhos. 2004. Global biodiversity informatics: setting the scene for a new world of ecological modeling. *Biodiversity Informatics* 1:1-13.
- Cushing JB, NM Nadkarni, B Bond and R. Dial. 2003. How trees and forests inform biodiversity and ecosystem informatics. *Computing in Science & Engineering* 5:32-43.
- Lin CC, JH Porter and SS Lu. 2006. A metadata-based framework for multilingual ecological information management. *Taiwan Journal Forest Science* 21(3):377-382.
- Michener WK. 2006. Meta-information concepts for ecological data management. *Ecological Informatics* 1(1):3-7.