

林業的無形資產—研究數據

- ◎林業試驗所森林保護組·林朝欽 (chin@tfri.gov.tw)
- ◎林業試驗所技術服務組·鄭美如
- ◎林業試驗所植物園組·陳建文

研究數據的重要性

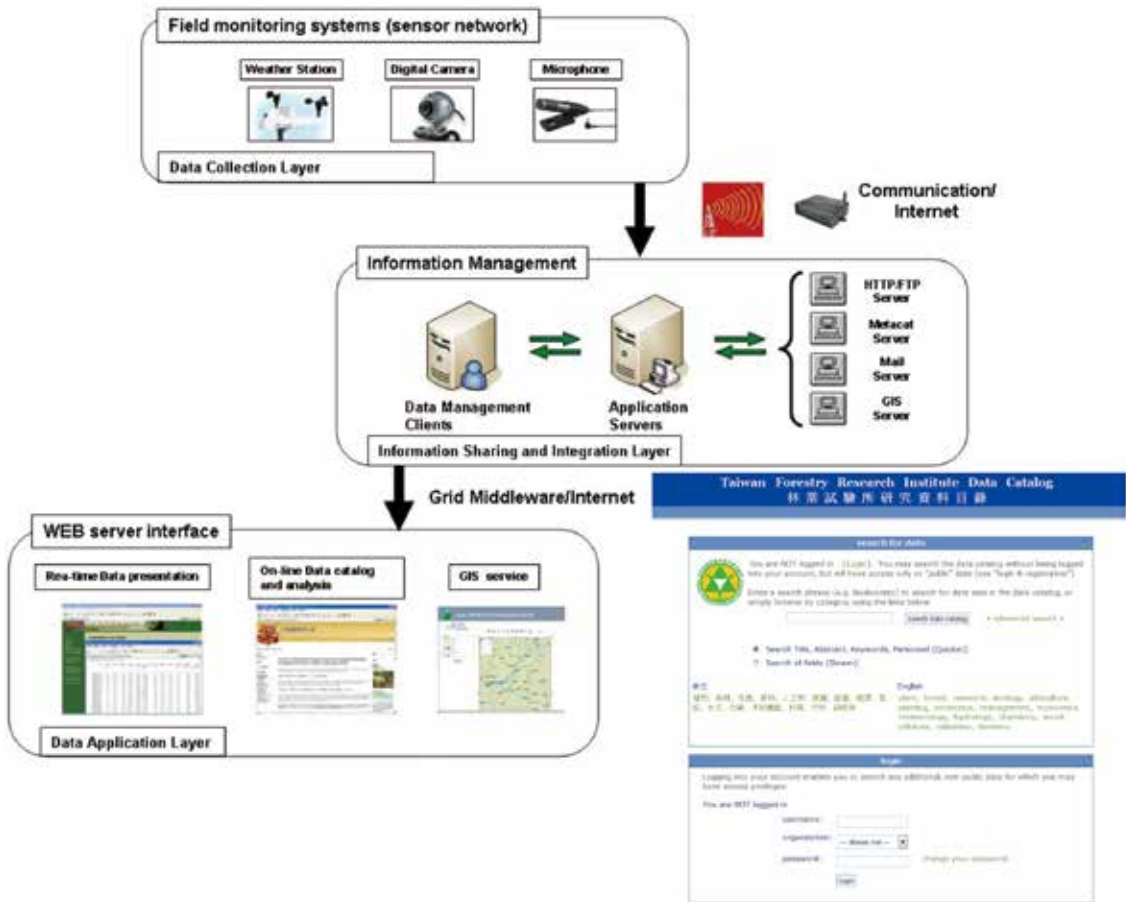
就長期生態研究觀點，沒有資料倉儲系統，資料會隨著計畫執行完畢而消失，另外其他各種新的研究問題、科學性資料的異質性、動態性、使用者是來自不同背景與目的等，均使得生態資料管理成為相當重要的問題。所以若目標是未來的20~50年後研究資料仍可取得且可說明解釋的話，要考慮的是資料政策、資料管理、倉儲、流通與整合技術性的方法、連貫性、與如何建立資訊管理系統的問題。就目前而言，一般的科學性資料多為研究者或研究計畫個別的將資料置於本身電腦上，沒有一定儲存標準格式，也沒有永久倉儲功能。不但難以分享，也不易加以整合再利用，所以為提昇資料本身之效益，資料之永久倉儲管理是有必要的，而欲達成此目的有賴建立一完善的生態資訊管理系統。然而生態資訊管理系統屬於資訊技術層面，在目前資訊技術發達的21世紀，並不難加以解決與克服，生態資訊管理系統能否建立與運作較難以解決的是資料管理政策的訂定、執行以及觀念的溝通，另外宣導與教育訓練也是很重要的關鍵。

生態資料一般泛指資源調查、分類學、生物學與生態學研究中所獲得的物種數據，最普遍的生態分布資料如：動、植物標本、森林資源調查的樹種名錄、自然保護區的物種普查等，是累積最多也最需要長期保存的原始數據(raw data or primary data)。更廣義的

生態資料可擴及與物種有關的環境、生態、生理與影像的資料。生態分布資料之所以重要，乃因此類資料具有時間、空間、分類群的資訊，在不同時間與空間尺度上提供了自然資源管理的重要基礎與變化的指標，尤其在全球氣候變遷的影響下，生態分布資料更是生物多樣性保育最重要的依據。

全球數據保存的共同趨勢

因此，全世界對生態資料的重視與管理均不遺餘力，例如美國的國家生物資訊基礎設施(National Biological Information Infrastructure, NBII)由其國家科學基金會補助設立完善的資料標準與管理政策，更架構資訊系統完善管理。又如全球生物多樣性資訊機構(Global Biodiversity Information Facility, GBIF)，於2001年成立，迄目前為止共有782個會員，累計蒐集生態分布資料達6.4億筆(其中有林業試驗所植物標本館103,172筆，及昆蟲標本館72,636筆分享到全世界)，可以說是全世界最大的生態分布管理單位，也是生態分布資料的管理核心組織。GBIF以完備的資料標準與資訊基礎架構管理與倉儲生態分布資料，已有許多科學研究使用GBIF所收錄的生態分布資料及進行物種分布預測，例如：哥倫比亞國際生物多樣性組織研究員Andy Jarvis的團隊利用GBIF的生態分布資料，運用生態區位模式及2050年氣候變遷模式預測66,000種植物的分布將移出南美各國所劃設的保護區，亦即現行的保護區範圍將無法保護



2006年林業試驗所建構的研究數據管理系統及林業試驗所的研究資料目錄。(http://metacat.tfri.gov.tw)

這些物種，提供了氣候變遷對物種保育影響重要的訊息。美國長期生態研究網(US LTER)規定各研究站在計畫結束後必須將研究數據倉儲在資料庫中，因此保存管理生態分布的原始資料是未來重要也是世界共同的趨勢。

資訊倉儲管理系統之建置

林業試驗所於2004年成立生態資訊研究小組，積極投入生態資訊管理資訊系統的研究，並開始建構研究數據倉儲系統，經過數年的投入，於2006年正式發布生態資訊

倉儲管理系統Metacat (Metadata catalog)，這個系統的架構以生態元資料語言(EML)為標準，利用網格(grid)技術為核心，屬於跨平台(Windows或Linux)的系統，且使用文字格式作為交換標準，讓機器與機器能互相傳輸資料。這個系統是利用已有的資料、及未來可能以自動化的無線網路儀器網把原始資料單獨處理，因此，不論資料以何種軟體建置均加以保存，如此可以免除生態學研究者重新處理資料的障礙。另外是系統對資料的品質管制與分析功能，此一功能由資料檢核及科

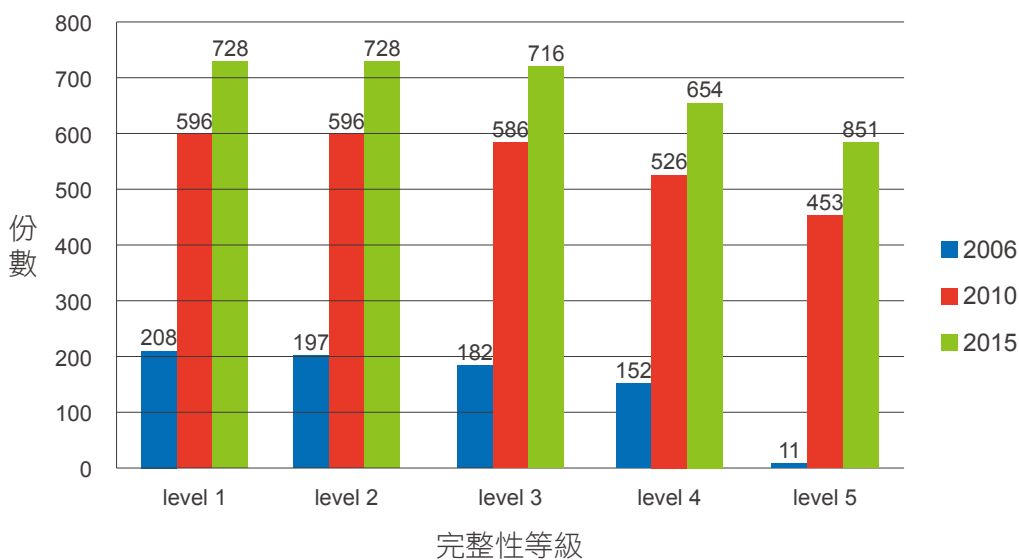
學工作流程(scientific workflow)來完成。科學工作流程其實是把科學家日常的工作過程加以規劃，並以標準的語法建立成數位檔案，讓系統能自動化完成資料分析與整合功能，且這種功能是可以重覆與分享的。

生態研究在長期的資料蒐集過程中，會累積數量龐大且差異性極大的資料檔案，加上資料檔案分散各地不易取得，使得即時獲取資料的分享機能無法發揮，不僅難以擴大資料的應用面，更因許多資料庫缺乏管理極易造成資料的流失，不但減損了這些研究資料的價值，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找到特定需要的資料；如何確保資料的長久保存與取得；以及如何確保資料的分享與整合，一直是生態研究人員所面臨的難題。為

了解決這些問題，需要建立一套標準，使得各系統之間有所遵循，才能進行資料的存取、索引、檢索與共享，元資料(metadata，即在一個標準化的格式下描述所蒐集到的資料)便是扮演這樣的角色與功能，因此元資料的推動與實施是非常重要的。元資料語言(Ecological Metadata Language, EML)便是一種根據生態學科而發展出的元資料規格。

生態元資料語言(EML)依其內容涵蓋範圍分成5個等級，第3等級即涵蓋所有研究相關訊息，但不包含原始數據的描述與實際的數據，而第4等級則包含原始數據的描述，若在加上原始數據就達到最高第5等級，這樣就完整的倉儲了該研究的資產了。為了協助生態學家能有一個容易使用與操作的介面，即使在不了解EML的規格、架構、各級標準等概念下，也能輕易的產生出一份EML文件，而開發出

EML文件完整性等級統計



林業試驗所Metacat中EML文件完整性等級。

了Morpho。Morpho是一種容易使用且跨平臺的資料管理應用工具，使得生態學家能產生、獲取和操作EML元資料和原始數據，並可查詢、觀看和檢索公開的生態學資料。

林業試驗所的研究資料目錄是生態元資料語言的第5等級的研究數據倉庫，2005年架設並中文化之後，於2006年開放使用，主要是林試所的研究同仁上傳研究資料，並以年終考績加分方式鼓勵同仁上傳過往的研究資料，第一年即獲得208份資料集，但含有原始數據的數量卻只有11份。次年，改變規定，發表的研究報告需上傳研究報告的metadata及原始數據方可計入年終考績分數，因此，至2010年所增加的資料集幾乎都含有原始數據，累積到453份。後因取消了此項強制規定，資料集的累積量就明顯下降。

因林試所的Metacat於2013正式成為地球觀測資料網絡(Data Observation Network for Earth, DataONE)的會員節點與Data Citation Index(DCI)的資料庫之一，為活絡資料庫內容，林試所於2015年又開始積極增納研究資料，規定研究同仁於所執行的科技計畫結案後，須於3個月內上傳資料，並通過資料審查小組的審核，以提升資料品質。截至2015年底，扣除因加入DataONE自動截取的資料集外，已累計728份資料集，而所增加的資料集也都含有原始資料。

林試所是臺灣極少數重視於研究資料倉儲管理的機構，推動這項工作的成效也相當不錯，在舉世開始倡導資料公開(Open Data)的時候，林務局於2011年也開始規劃訂定委託計畫成果的資料繳交，在林試所協助下訂定了「林務局生態分布調查研究計畫資料繳交與

使用作業說明」，並建置了「林務局生態調查資料庫」(網址：<http://metacat.forest.gov.tw>)。資料繳交與使用作業說明中明定委託計畫必須於結案前完成資料繳交，截至2015年底，完成了98個委託計畫的資料繳交與公開，加上2008年至2011年補建置上傳的資料集，共累積了285份，原始數據已累積超過50萬筆。因為林試所與林務局Metacat都使用共同的標準，所以可以透過網路服務(web services)功能互相交換與共享，這兩個研究資料倉儲系統將成為林業研究上很重要的無形資產。

由林試所與林務局的經驗看來，生態資訊管理系統能否建立與運作，當觀念尚未普及或被接受時，要保存這些寶貴的研究資料，達成資料永久倉儲，最重要的因素在於資料管理政策，由上往下推動。在全球推行開放資料的風潮下，冀望林業的研究資料可以不因個人的研究領域轉換或其它因素而流失，可以將這些研究資料有系統的進行倉儲，讓這些無形的資產可以延續與分享。⊕



林務局的生態調查資料庫。