



研究報告

生態資料管理與分享的新嘗試：語意連結

林朝欽^{1,3}，麥館碩²，夏禹九²

¹林業試驗所森林保護組；²國立東華大學自然資源管理研究所；³通訊作者 E-mail: chin@tfri.gov.tw

[摘要] 生態學是一門重視領域合作與資料分享整合的科學，但傳統的生態研究並未達到這樣的境界。本研究嘗試以鏈連資料技術，以實作的方式，將現有生態資料庫加以語意連結。本研究主要目的在展示鏈連資料技術有別於現行元數據或關連式資料庫之生態資料管理與分享框架，以達成較理想的生態資料管理、整合與分享需求。本研究使用了森林火災、植物標本、昆蟲採集、森林動態樣區調查以及台灣的物種名錄等五個資料庫，以資源描述框架作為資料標準連結資料，讓原本分散難以分享的資料在最細的單元彼此連結，形成有別於傳統網路的資料管理與分享方式，並讓機器協助資料處理作業，跨越了系統的界線的連結限制。本研究將鏈連的內容命名為鏈連開放的生態資料，在 <http://ecowlim.tfri.gov.tw/lode> 上提供研究者使用。

關鍵字：資源描述框架、鏈連資料、生態學、資料分享。

Ecological Data Management and Sharing Using Semantic Link: A New Approach

Chau-Chin Lin^{1,3}, Guan-Shuo Mai² and Yue-Joe Hsia²

¹Division of Forest Protection, Taiwan Forestry Research Institute; ²Department of Natural Resources and Environment Studies, National Dong Hwa University; ³Corresponding author E-mail: chin@tfri.gov.tw

ABSTRACT Ecology is a discipline emphasizing integrative and collaborative approaches. In terms of data management and sharing it hasn't reached that state. In order to facilitate the integration of ecological data, we attempt to build a new approach by using Open Linked Data technique. The main purpose of our study is to demonstrate the difference of this approach from metadata and relational database methods of data management and sharing adapted currently. The practice uses Resource Description Framework (RDF) to make data linkable and machine-readable in the finest level of granularity, interweaving data silos into the Web of Data. The new-proposed framework for sharing ecological data is so named Linked Open Data of Ecology (LODE). There are 5 datasets including forest fire database, plant and insect specimen databases, the Nanshi forest dynamics plot census and the catalogue of life in Taiwan tested in this practice. To show the scalability and interoperability, the web [http://ecowlim .tfri.gov. tw/lode](http://ecowlim.tfri.gov.tw/lode) is set for use.

Keywords: RDF, Linked Data, Ecology, Data Sharing.



前言

生態學是一門重視跨領域合作與資料分享整合的科學，但傳統的生態研究目前並未達到這樣的境界，傳統生態研究一直以個人或小團隊主導的、小尺度時空的、短期性的觀察與實驗為主流（夏禹九等 2000）。這樣的研究模式雖累積了數量龐大的資料，但差異性極大，這些資料大部分由個別科學家或研究室管理，因而分散各地，且以不同的方式與工具管理資料，使得資料的流通與分享困難，其他科學家往往不知道有那些研究資料，也無法取得這類資料，更談不上整合資料（林朝欽等 2006）。

1970 年代開始，由於個人電腦的出現，研究資料開始進入數位化格式，許多複雜的生態現象藉由數位化的資料分享與整合找到新答案，此利器促使生態研究的模式發生改變。自 1980 年代開始，跨領域團隊的、大尺度時空的、長期性的研究成為新主流（Reichman *et al.* 2011）。此演化進一步導引生態研究逐漸成為以密集性與多面向資料探索與整合為主的科學，對於資料分享的需求因此愈發顯著，因為傳統的研究模式已無法收集與整合足以解決較大的生態問題的資料（Porter 2010）；更甚者傳統生態研究難以分享已發表文獻背後的原始資料，對強調「可重複」驗證生態假說的理想無法達成（Parr and Cummings 2005）。雖然資料分享的需求日益增加，即便各研究學門皆廣為強調，但資料分享在實務上仍未普及（Anonymous 2009）。其原因被認為是推動資料分享仍存在著技術與文化上的藩籬（Parr and Cummings 2005）。

近年來部份國家或國際間對於網路以及資訊科技的投資，使得精良的資訊軟硬體設施系統得以建置，大大降低資料分享的技術性困難，加上科學期刊改變出版政策，新的生態資料分享文化在生態研究領域已逐漸形成，而使得生態研究正邁入一個新的研究典範（Arzberger *et al.* 2004, Porter 2010, Reichman *et*

al. 2011）。

生態資料分享在網路上的技術可以突破拜語意網（Semantic Web）（Berners-Lee *et al.* 2001）技術開發之賜，語意網相較於現在普遍的全球資訊網（World Wide Web, WWW），是一個更為理想化的資料網絡。全球資訊網主要是以「文件」為連結單元的網絡；而語意網則是透過賦與事物以及關係特定的名稱與型別來跨越至解析度更高的「資料」網絡。語意網不僅被設計為一新的資料模型，它特別適合於連結在許多不同資料模型之下的資料，允許全球資訊網上的資料庫或文件加上額外的資訊描述（額外的一層語意），以及允許跨資料庫的複雜操作（Berners-Lee *et al.* 2001）。隨著語意網技術的日趨成熟，對協助生態資料被發現、取得以及使用具有相當大的應用潛力。尤其利用語意網概念中的鏈連資料（Linked Data）技術，可以發布與相互連結全球資訊網上結構化的與分散的資料供分享與再使用，此技術在各個領域中正日益蓬勃發展（Bizer *et al.* 2009）。在生物相關領域中，雖然尚未普遍，但已有與物種多樣性相關的資料集以鏈連資料的形式發布上網，全球生物多樣性資訊機構（Global Biodiversity Information Facility, GBIF）自 2009 年起也正籌措生物多樣性資料的發布計畫。

所謂鏈連資料是一個語意網脈絡下的概念與資料分享風格，它依循 4 個原則建立可以分享的資料內容（Berners-Lee 2006）：1. 使用統一資源標識（Uniform Resource Identifier, URI）當作事物的名字；2. 使用超文件傳輸協定（HyperText Transfer Protocol, HTTP）的 URIs 來使人們得以查循這些資源；3. 當使用者找到某一個 URI，使用特定標準來提供使用者有用的資訊；4. 使用者所獲得的資訊包括了通往其他 URIs 的連結，讓使用者可以找到更多事物。鏈連資料提供了一個資料發布與分享的新典範（Bizer *et al.* 2009）—亦即資料將直接在全球資訊網上彼此連結，並以全球尺度的資料空間呈現，這個全球資料空間可以在全球資訊網上更精確地描述真實世界中各事物的特質



林朝欽，麥館碩，夏禹九

與對應關係，因此可以稱之為資料網(Web Of Data)。各連結的資料更進一步可交互連結成為龐大的鍵連開放資料雲 (Linked Open Data cloud)(Bizer *et al.* 2007)。迄 2010 年底已有超過 200 個資料集鍵連開放資料雲上，使用者能夠輕易的依循語意連結流覽已整合的資料集，如同一般使用者流覽網頁上的文字或影音(Jain *et al.* 2009)。鍵連資料也提供了許多電腦可以自動處理資料的機制，利用資料內容的語意連結，此機制讓機器得以在極少人力介入下自動發現並整理全球資訊網上的資料，便利需要大規模彙整資料的研究(Reichman *et al.* 2011)。

本研究嘗試以鍵連資料技術，透過實作的方式，將現有生態資料庫加以鏈結，本研究主要目的在展示鍵連資料技術有別於現行元數據或關連式資料庫之生態資料管理與分享框架，達成較理想的生態資料管理、整合與分享需求。本研究共使用了五個資料庫，將之鍵連，並把鍵連後可使用的整合內容命名為鍵連開放的生態資料(Linked Open Data of Ecology, LOD E)，呈現在現有的資料網上提供生態研究者使用。

材料與方法

本研究使用五個不同來源且分散的資料庫，分別是林試所的林火生態學資料庫(Firedb)、植物標本館資料庫(TAIF)、昆蟲標本館資料庫(Flyhorse)、林務局委託靜宜大學的楠溪森林動態樣區調查資料庫(FDP-NS)、以及中央研究院生物多樣性研究中心的台灣物種名錄資料庫(TaiBNET)。林火資料庫倉儲 1957 年至 2009 年間的 2600 筆林火資料；昆蟲標本資料庫存有 11 萬餘筆標本資料；植物標本資料庫具有 10 萬餘筆植物標本資料；楠溪森林動態樣區調查資料為 2005 年第一次調查內容，含樣區內 18790 株之木本植物資料；台灣物種名錄資料庫由中研院生物多樣性研究中心建置，為向分類學家徵求而來的台灣各類群物種

之名錄資料，累計有 53253 種不同類群物種名錄清單，及詳細之分類層級資料與學名的使用現況。

資料集的建立與發布流程可歸納為四個工作項目：1. 準備原始資料並建立資料模型；2. 將資料轉換為 RDF 的形式；3. 將 RDF 資料載入專屬的系統，並以推理器建立資料中隱含的語意；4. 透過鍵連資料介面發布資料並與其他資料網上的資料集建立連結。整個工作流程：首先對資料內容進行分析，以找出它們涵蓋的主題以及相關性，資料進行轉換前尚需要進行檢核與校正，以使機器得以自動操作資料轉換的工作，而後建立統一的資料模型，並挑選、重複使用合適的字彙來增進與資料網的交互操作性；其次是將非 RDF 之資料(包括了五個關連式資料庫與生態元數據(Ecological Metadata Language, EML)文件轉換為 RDF 格式，使資料與 EML 文件得以在單一框架之下被描述與使用。最後透過鍵連資料界面的應用程式，將資料正式以鍵連資料的形式發布上網提供連結，並使機器能隨著具備語意的 RDF 連結運行於資料網中找尋相關資訊。

結果

一、資料庫內容分析

林火資料庫的資料量 1Mb，內容主要分為 8 個資料表，以 Fireinfo 資料表為中心，內含 2600 筆林火資料，時間尺度橫跨 1957 年至 2009 年。Fireinfo 資料表記錄了林火的發生時間、發現時間、受控制的時間與撲滅時間、動用的消防隊員額、以某一特定記錄年份為界，先後使用大地座標系統 TWD67-TM2 或 TWD97-TM2 的地理座標位置、以及林火影響面積等資料。向外延伸，其它資料表記錄了林火發生地包括林業經營管理觀點的林班地、事業區、林區管理處等資訊；另有採行政觀點的林火發生地如各級縣市的資料。資料品質上具有如下的問題：1960 年之前的史料並無確實記載林火位置，而是以林火發生之林區中心點



座標為代表、林火發生時間資料中有許多可能為輸入時順序顛倒造成林火發生時間點晚於林火撲滅時間點的錯置情況，因此與林火相關之時間資料，僅林火發生之時間點較具使用價值。

昆蟲標本資料庫的資料量 40Mb，內容共分 33 個資料表，以昆蟲標本資料表為核心，記錄 11 萬筆標本資料及標本的鑑定、歸檔、保存等日期以及以度分秒制之座標，但並未記載所使用的座標系統。其它資料表記有外出採集標本事件，採集地點(國家、縣市、鄉鎮)，標本採集者，標本鑑定者，分類資料(界、門、綱、目、科、屬、種)，標本製作方式，現保存位置(標本櫃)，以及模式標本等資訊。本資料庫中部分資料表在資料庫更動的歷史中已廢棄無實際使用，此外有部分涉及標本出借或資料庫管理身份等涉及隱私之資料。另外資料品質有如下的數項問題：資料具有多餘的空白或亂碼符號、人物或地點描述方式或格式不一致，例如「台北縣」、「臺北縣」或「Taipei County」、「Taibei Co.」等地名上的異名用法、中文人名轉換為英文拼音後的姓與名順序、拼法、縮寫格式不同、部份屬於欄位與資料不一致的狀況，例如內容按照不同語言(中文、英語)而區分的欄位，當中卻有部分未按照規則填寫對應語言之資料而出現中英夾雜、未被鑑定的標本在物種欄位中填入科或屬的資料、以及名為 county 但內容填入縣市資料。

植物標本資料庫的資料量 174Mb，內容僅一個資料表，113 個欄位，共 10 萬筆資料，約為完整的林業試驗所植物標本資料的一半數量。資料表的欄位名稱多半依據 GBIF 採用的 Darwin Core 1.4 版標準而制定，目的是利於搭配分類學資料庫工作組(Taxonomic Databases Working Group, TDWG)所開發的 TDWG 資訊收集存取協定(TDWG Access Protocol for Information Retrieval, TAPIR)及其工具，以與 GBIF 進行全球生物多樣性資料的彙整。內容涵蓋標本的發現記錄，標本類型，採集事件、保存地點與分類等相關資訊。此外尚有部份欄

位內容是參考 Darwin Core 而自訂的欄位，多半設計為專門存放以中文描述之資料，做為雙語系使用。資料庫中的資料品質問題如同昆蟲標本館資料庫，但因其中所儲存者已是經過程式處理轉換後的資料，因此資料格式與內容較為一致。由於本資料庫設計之初並未以標本採集事件做為採集的時間或地點的記錄單位，而單一件標本採集日期必定只有一日，因此 Darwin Core 中雖有要求分別記錄採集開始與結束時間，但本資料庫中多數情況下二者資料是重覆的，而例外情況通常是缺少記錄採集結束的時間之故。單件標本的採集地點、採集者與鑑定、命名者，很多時候在資料庫中會出現一欄多值的情況。由於過去建立資料檔案的習慣上視人為主要使用對象，未將此類資料先行正規化。

靜宜大學楠溪森林動態樣區調查原始資料最初是以逗號分隔數值(Comma -Separated Values, CSV)格式儲存，內容涵蓋了 2005 年第一次動態樣區調查樣區中總計 18790 株、胸高直徑大於 1 cm 之本木植物的 32493 個分枝的量測資料。這些資料經由林業試驗所資訊工作小組成員建立 EML 文件，描述了調查計畫的概述、調查時間、樣區地點與西南角座標、調查人員以及存放調查結果的資料表中的欄位與變數定義，包括樹木與分枝的編碼、樹種、每棵樹在小樣方或樣區中的相對位置等，並匯入由內容管理系統(Content Management System, CMS)所架設之網站中。

台灣物種名錄資料庫是由國科會贊助之計畫，並由中研院與農委會合作建置，並向台灣的生物學專家徵求而來的台灣各類群物種之分類資料以及名錄資料，本研究進行時，資料庫中累計有 53253 種不同類群物種，堪稱台灣現存單一資料庫中最為完整的物種名錄與分類資料。資料庫中主要分為 17 個資料表，資料量近 90Mb，當中除記有物種名錄清單，亦有詳細之分類層級資料與學名的使用現況。資料內容中有部分分類不明確或學名拼法有所疑義者，偶會以英文句號、中括號、問號

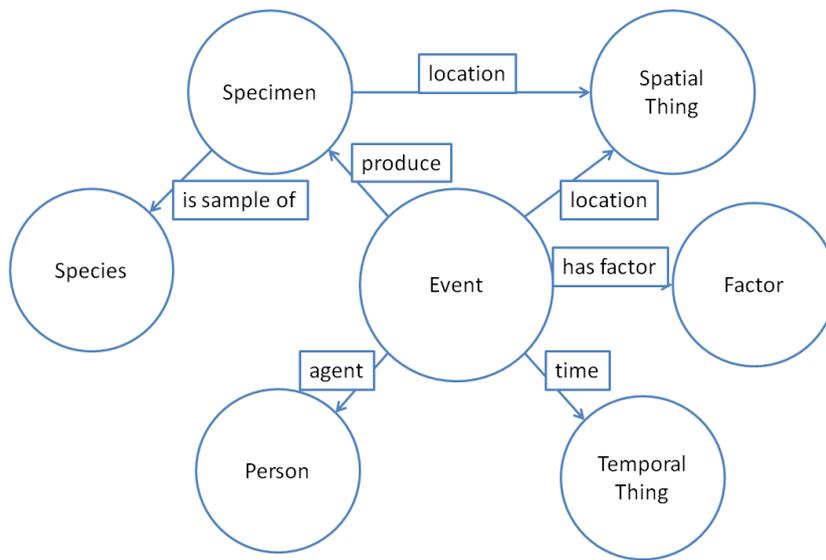
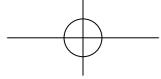


圖 1. 以事件(event)做為中心，各資料庫可描述與事件相關的其他種種資訊，包括了事件的時間(temporal thing)、地點(spatial thing)、原因(factor)、參與者(person)、事件產出的事物(如標本 specimen 資料)等；產出的事物則另外可與物種分類的概念(species)相連接，做為事件與分類的橋樑

等記號做為註解或表示保留之意；資料庫中部份中文字疑似因編碼問題，成為可能引起系統錯誤的亂碼。

二、 資料庫內容之語意連結

由各資料庫內容分析所得，可以歸納出幾個主題概念與關係，分為：事件(研究計畫、林火、標本採集、量測)、發生的原因、事件發生與結束的時間、地理資訊如事件發生地點(國別、行政區、林業管理分區、座標)、參與事件的人物、事件中產生的事物或關注的對象(標本、樹木主幹與分枝)、以及與這些事物內在本質相關的外加資訊(如將生物依特徵歸類後給予的物種的學名與分類的階層等等)，關係整理如圖 1。圖 1 所示僅為普遍化概念之呈現，各概念可再細分描述更為精確的子類別或子屬性，如空間屬性的事物(Spatial Thing)類別中有國家、縣市等跟地點相關的子類別，人物可再分為標本採集者、量測者、計畫主持人等等。

物種的概念在本研究中，是經由對實體的觀察歸類命名而成，而我們可以視標本為一個

已被社群認可的物種的樣本(samples)。一物種的分布地是經由觀察者在該處採集或調查事件中發現到的個體或採集到的樣本間接建立起物種這個概念與地點間的關連。因此研究者(例如分類學者)只需要維護某一標本與其被認定的物種「身份」(identity)，即可一併維持物種與分布地點資料的一致性；而在對物種的分類狀況有歧見時各資料擁有者也能藉由每件標本都是獨一無二的特性，保有研究者對物種身份的認定。但在上述的關係圖中仍缺少了可以描述生態研究中最普遍的、記載有各種量測值的資料表。參考 EML 文件中對資料表的描述項目，本研究就「事件」的子類別「量測事件」建立另外一個足以描述多數資料表的關連模型(圖 2)。這個模型對於量測事件的定義實際上並非以「進行一次純粹的量測動作」當作定義，而是以「量測者針對單一對象進行的一系列量測行為，這些行為以現有資料為參考，共享有不可再細分的時間或空間實體的脈絡」當作一次事件的準則，因此在一次量測中可能得到多個量測值，圖 3 是楠溪森林動態樣區調查計畫所觀測的實體、原始觀測資料、轉換

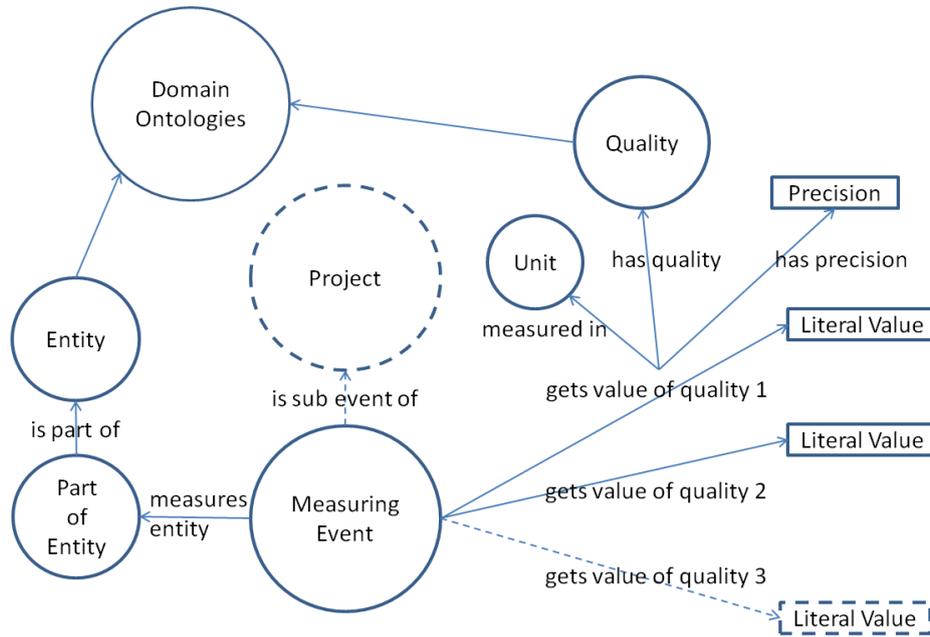


圖 2. 依 FDP-NS 資料集的 EML 文件中描述所建立的資料模型；每一次的量測事件(measuring event)都會自其量測對象身上得到(gets value)一個或數個量測值(literal value)；這些量測值各自代表量測對象(entity)的某些特徵(quality)，並使用了慣用的單位(unit)來描述這些值

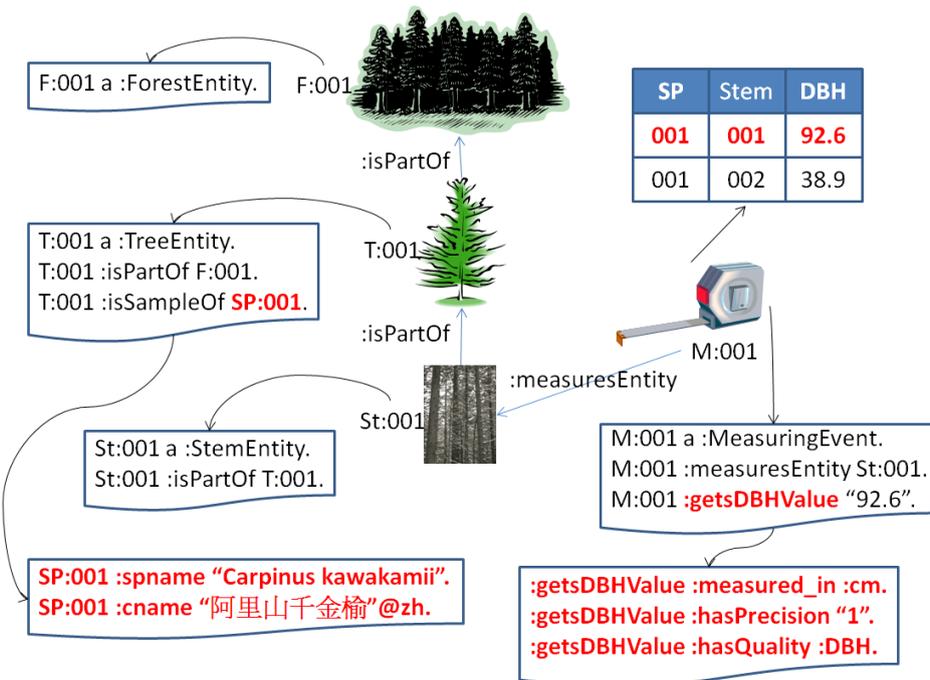


圖 3. 楠溪森林動態樣區調查計畫所觀測的實體、原始觀測資料、轉換後的 RDF 資料以及元數據之間的關係示意

TAIF:PLANT:135034:1:1:219029 at Linked Open Data of Ecology
<http://ecowlim.tfri.gov.tw/lode/resource/taif/Specimen/TAIF-PLANT-1-1-135034-219029>

| Property | Value |
|------------------------|--|
| dwc:basisOfRecord | PreservedSpecimen (en) |
| dwc:catalogNumber | 219029 (en) |
| taif:collectedInEvent | taif:CollectingEvent/ShengYouLu19860207Lienhuachih19860207 |
| eco:collectedInEvent | taif:CollectingEvent/ShengYouLu19860207Lienhuachih19860207 |
| dwc:collectionCode | PLANT (en) |
| dwc:country | Taiwan (en) |
| dwc:county | Nantou County (en) |
| eco:lit_dateCollected | 1986-02-07T00:00:00 (xsd:dateTime) |
| taif:lit_dateCollected | 1986-02-07T00:00:00 (xsd:dateTime) |
| dwc:dateIdentified | 1986-02-07 (xsd:date) |
| dwc:disposition | in collection (en) |
| taif:lit_duplID | 1 (xsd:integer) |
| eco:lit_duplID | 1 (xsd:integer) |
| is:event:hasProduct of | taif:CollectingEvent/ShengYouLu19860207Lienhuachih19860207 |
| is:taif:hasSample of | taibnet:Species/Smlax_sp. taif:Species/Smlax_sp. |
| is:eco:hasSample of | taibnet:Species/Smlax_sp. taif:Species/Smlax_sp. |
| dwc:higherGeography | Taiwan, Taiwan, Nantou County (en) |
| eco:lit_id | 1 (xsd:integer) 135034 (en) |
| eco:identifiedBy | taif:Person/ShengYouLu |
| taif:identifiedBy | taif:Person/ShengYouLu |
| dwc:identifiedBy | Sheng-You Lu (en) |
| dwc:institutionCode | TAIF (en) |
| eco:isSampleOfSpecies | taibnet:Species/Smlax_sp. |
| dwc:higherGeography | Taiwan, Taiwan, Nantou County (en) |
| eco:lit_id | 1 (xsd:integer) 135034 (en) |
| eco:identifiedBy | taif:Person/ShengYouLu |
| taif:identifiedBy | taif:Person/ShengYouLu |
| dwc:identifiedBy | Sheng-You Lu (en) |
| dwc:institutionCode | TAIF (en) |
| eco:isSampleOfSpecies | taibnet:Species/Smlax_sp. |



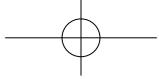
圖 4. 直接瀏覽鍵連資料界面中的標本資料，圖中可見與標本相關的資訊以及標本之數位化照片

Find a species resource from Linked Open Data of Ecology by any text:
Note : This will take longer time than searching by scientific name and will return only first 100 results.
Please input any text here:

Find a species resource from Linked Open Data of Ecology by scientific name:
Please input genus name here:
and/or specific epithet here:

您也可以直接輸入一個resource (請連同角括弧一併輸入):
範例: <http://ecowlim.tfri.gov.tw/lode/resource/taibnet/Species/Lonicera_japonica_chinensis>
<<http://ecowlim.tfri.gov.tw/>>
choose your data source:
from Web Of Data
from LODE (local)

圖 5. 物種名與分布地查詢操作頁面，畫面上列是接受任意關鍵字的全文查詢，中間一列是以物種學名(可單用屬名或屬加種小名)查詢，最下方可由已知的 URI 直接取回相關資源



後的 RDF 資料以及元數據之間的關係，圖中以 URI 為觀測實體命名並以 RDF/N3 描述了一森林(F:001)、樹木(T:001)、樹幹(ST:001)與量測事件(M:001)與其量測值之間的關係。自一 URI 出發的箭頭指向了存有與該 URI 相關資料之 RDF 文件。物種(SP:001)資訊與量測使用之標準(如單位、精確度與物理量)等較為抽象的元數據概念，另也額外描述了資料欄位內容各自代表的意思、欄位間的關係以及其與觀測實體之間進一步的關係。

三、資料庫內容之發布與使用

透過提供鍵連資料界面 <http://ecowlim.tfri.gov.tw/lode>，本研究已把資料正式以鍵連資料的形式發布上網，提供可連結、查詢，使用者能順著具備語意的連結運行於資料網中找尋相關資訊。此外系統尚可主動與其他資料集做連結，讓使用者獲得更多的資訊。換句話說，在鍵連資料的世界裡，「當你擁有一些資料，你就能找到更多」。本研究以下列的三個實例說明以鍵連資料的方式發布的資料集如何提供使用者查詢資料內容，以及利用鍵連關係與其他資料集之間的連結，將本研究之鍵連資料界面視為入口(portal)，使用者可以介接鍵連開放資料雲，使用與生態研究相關的資料集。這些資料可直接供人瀏覽內容，抑或是經由機器處理消化後，由機器將彙整的結果呈現於使用者面前。

1. 直接瀏覽

使用一般的網頁瀏覽器，使用者可以無阻礙地在本研究已發布的資料集中查詢整合性內容，亦可與其他相連的資料集之間進行跨資料集的查詢，並瀏覽或取得原始資料，並由連結對象的主題、內容與組成的方式獲得這些資料的背景脈絡，使用者可參考屬性描述之資訊，或是對資料集的描述之資源，找到本資料集主要之資料類別，例如一份標本的分類、採集、地點、人物、時間等。若再依循這些資訊的連結，使用者可發現更多相關之資訊，圖 4 為直接瀏覽鍵連資料界面中的標本資料，圖中

可見與標本相關的資訊以及標本之數位化照片。

2. 物種名與分布地查詢

使用者可以用學名查詢，並直接瀏覽物種之資料。以忍冬(*Lonicera Japonica*)這種植物為例，與忍冬這個物種有關之資源及名稱分別有 8 個資源以及 12 個名稱，而分布地的資源則有 88 個。名稱如 Japanese Honeysuckle、Slingertry、Chèvrefeuille du Japon 與スイカズラ等不同的俗名，分布地則可藉由已鍵連的 GeoSpecies 資料集，找出了在本研究的資料集中未記錄的分布點，本例中查詢到忍冬也分布在美國的威斯康辛(Wisconsin)、密西根(Michigan)與麻薩諸塞(Massachusetts)三大州之中(圖 5)。

3. 資料間的關係探索

本研究資料集間的相互關係可以透過關鍵詞來探索，使用者藉由關鍵詞可以探索語意連結關係而形成一個關係網，圖 6 以「烏鴉鳳蝶」與「南投縣」兩個關鍵詞為例。系統以動態圖形呈現烏鴉鳳蝶與南投縣之間的關連。結果除了物種與地點的直接關連之外(烏鴉鳳蝶於南投縣有分布)，亦找出了同在南投分布的同屬蝶類(寬青帶鳳蝶、昇天鳳蝶)、在南投採集過烏鴉鳳蝶的學者 Y. C. Chang 同樣在南投參與過的其他採集事件，以及烏鴉鳳蝶在另一個與南投縣相關的資源—台灣—與台灣其他縣市的分布情況。鍵連資料的語意連結特性讓我們得以此工具提供的方法探索廣大資料中，任意概念之間的關係。

討論

生態學研究資料的整合除了依循相關理論與框架(Pickett *et al.* 2007)外，尚可依詢現有資料的語意關連來達成，但目前按計畫或研究目的賦與資料單一向與定位的資料管理方式，以及以文件做為資料分享的單元，顯然難以支持生態學的情境模擬或資料探索的典範。本研究利用鍵連資料進行資料的語意連結

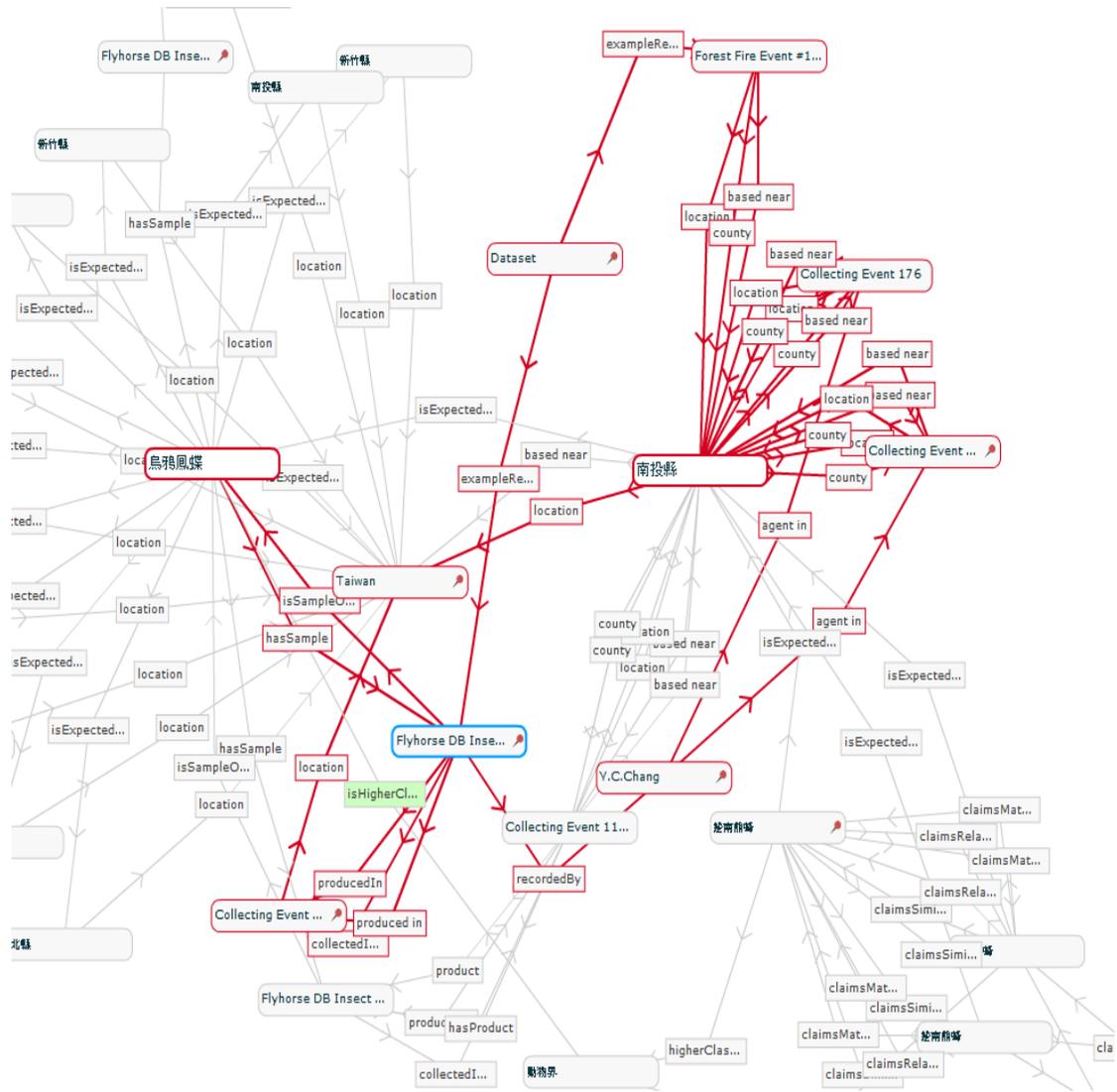
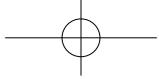


圖 6. 本圖中展示了烏鴉鳳蝶與南投縣之間的關連性；除了由最基本的標本與採集事件而連結，還包括了其他分類群在南投的記錄關係，或是由同在南投但屬不同採集事件，由同一記錄者串連起的關係等

，其尺度可變動性高，例如本研究的五個資料集，可將計畫、資料集、研究對象彼此關連，達到多層面的資料探索功能。在鍵連資料框架中，我們可依現階段對生態實體、現象與層級關係的認知來組織觀測資料，除了可依不同問題或子領域的知識模型為脈絡突顯一實體在不同面向所處的焦點地位，亦能藉由相連的資料間接做為生態實體與現象在特定觀點與時空脈絡下的特徵代表。雖然本研究所測試的鍵連資料技術，並不是生態學整合問題的完全解答，但它的語意連結為生態學的資料探索典範

提供了合適的資料與知識的描述框架。這些物件化的概念毋須如表格文件、資料庫系統等，在倉儲前就先被強制定位、給予固定的屬性或儲存空間。將分屬不同子範疇，但與同一實體相關之資料，依我們現有的、描述該實體的知識模型來組織連結，讓使用者能容易地理解與取得這些資料，進而能在日後有需要時，能在第一時間幫助研究者驗證其假說，抑或是完備其理論構思。既然學界已證實了使用以計畫為標的，摘要與歸納出資料集內容的 EML 標準能夠有效地協助生態資料管理，本研究認為以



鍵連資料技術管理生態資料將是個更為合適的選擇。

鍵連資料技術除了提供一個較新的生態資料管理，更提供了後續的資料整合與分析的便利性。我們將某些特定而繁複的資料查詢動作重現於網路，並將之轉換成後續將使用的特定整合或分析功能。亦即具有語意查詢的資料服務可以輕易地整合在後續的資料分析流程中，來實現資料的共享。

本研究在發布資料之前，面臨了資料庫的資料品質的問題，雖然這些關連式資料庫的欄位名稱、資料型別的記錄具備了某些程度的語意，每個欄位中的資料項目仍會因格式、語法、慣用語或對欄位定義認知不同等因素造成資料語意上難以分辨或不一致的情況。其原因為填寫資料內容時未具標準化，例如標本資料，在填寫當下無從得知或鑑別較低分類階層，硬填入較高層的資訊，但更適合的方式是以留白或標記以特殊記號的方式來處理；採集日期無標準格式等，因此建立簡單的控制字彙 (controlled vocabulary)，是改善生態資料品質應考慮的方式。

本研究處理物種學名以「所有分類層級中的名稱都是獨一無二」的前提對應，但結果發現單一的關係完全不足以描述看似簡單的概念間的對等關係，例如「*Satyrium*」同時為植物資料集中蘭科 (*Orchidaceae*)、與昆蟲資料集中小灰蝶科 (*Lycaenidae*) 下的一個屬 (genus)；「*Axiidae*」則同為昆蟲資料及中鱗翅目 (*Lepidoptera*) 與十足目 (*Decapoda*) 下的一科 (family)。顯然在生態資料中涉及分類學的身份對應的表示方法，有待更詳細的探究，尤其是單一的關係蘊含的特性一但被實作於系統中，便會發生一件事物得到不屬於它的屬性的連結錯誤 (McCusker and McGuinness 2010)。因此，如任一概念具有隨著時間變化的特性，例如縣市名稱、轄區界線、分類學名、一地之生物個體數、一地之人口等，我們需要結合能夠描述事物隨時間沿革的字彙集，來因應不同時空背景產生的不一致性。

結論

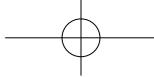
以生態學的歧異與複雜性來說，生態資料管理難有單一合適的資料管理方式。應用元數據標準為導向的管理方式雖保留資料的原始性質，但缺乏足夠的語意與連結，在鍵連資料的框架下，以 RDF 描述的資料可以在避免使用「標準組合」的情況下，記錄更多可能對日後重要而有用的訊息，也能解決因不同目的而建立的資料儲存、存取與描述標準上的差異。本研究共使用了森林火災、植物標本、昆蟲採集、森林動態樣區調查以及台灣的物種名錄等五個資料庫，以資源描述框架為資料標準，讓資料在最細的單元彼此連結，形成有別於傳統網路的資料網，並能讓機器協助資料處理作業，跨越了系統的界線而彼此連結。

誌謝

本研究之完成有賴林業試驗所林火研究室、植物與昆蟲標本館、靜宜大學生態研究室提供測試之資料集、及中研院生物多樣性研究中心建置且公開分享的物種名錄，作者謹此誌謝。另林業試驗所生態資訊研究小組之資源與技術支援亦為本研究完成之重要關鍵。

引用文獻

- 林朝欽、陸聲山、鄭美如。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
- 夏禹九、金恆鏞、林敏雄。2000。長期生態研究的台灣經驗。科學發展月刊 28(9): 679 - 685。
- Anonymous. 2009. Data's shameful neglect. *Nature* 461(7261):145-145.
- Arzberger P, P Schroeder, A Beaulieu, G Bowker, K Casey, L Laaksonen, D Moorman, P Uhlir and P Wouters. 2004. An International Framework to Promote Access to Data. *Science* 303(5665):1777-1778.
- Berners-Lee T. 2006. Linked Data. *International*



林朝欽，麥館碩，夏禹九

- Journal on Semantic Web and Information Systems* 4(2):1. Available at: <http://www.w3.org/DesignIssues/LinkedData.html>.
- Berners-Lee T, J Hendler, and O Lassila. 2001. *The semantic web*. *Scientific American* 284(5):34-43.
- Bizer C, T Heath, D Ayers and Y Raimond. 2007. Interlinking Open Data on the Web. Pages 802-815 *In* the 4th European Semantic Web Conference.
- Bizer C, T Heath and T Berners-Lee. 2009. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems* 5(3):1-22.
- McCusker JP and DL McGuinness. 2010. Towards Identity in Linked Data. *In* Proceedings of OWL Experiences and Directions Seventh Annual Workshop.
- Parr CS and MP Cummings. 2005. Data Sharing in Ecology and Evolution. *Trends in Ecology and Evolution* 20(7):362-363.
- Pickett STA, J Kolasa and CG Jones. 2007. *Ecological Understanding: the Nature of Theory and the Theory of Nature*. 2nd edition. Academic Press, Waltham, Massachusetts, United States.
- Porter JH. 2010. A Brief History of Data Sharing in the U.S. Long Term Ecological Research Network. *Bulletin of the Ecological Society of America* 91(1):14-20.
- Reichman OJ, MB Jones and MP Schildhauer. 2011. Challenges and Opportunities of Open Data in Ecology. *Science* 331(6018):703-705.