

語意網技術在館藏目錄中的應用：以原住民文物典藏目錄資料為例

The Application of Semantic Web Technology in
Collection Catalog: Take the Cultural Collection Catalog
of Indigenous People as an Example

戚玉樑

Yu-Liang Chi

中原大學資訊管理學系教授

Professor

Department of Information Management

Chung Yuan Christian University

丁嬋娟

Chan-Chuan Ting

中原大學會計學系講師

中原大學商學博士學位學程博士生

Instructor, Department of Accounting,

Chung Yuan Christian University

Ph.D. Student, Ph.D. Program in Business,

Chung Yuan Christian University

【摘要 Abstract】

近年來，語意網（Semantic Web）已從概念進入到實現階段，隨著相關技術和軟體工具發展成熟，實務應用更需要機構（內容所有者）積極參與，特別是提供可操作的資料。由於鏈結開放資料（Linked Open Data, LOD）允許大眾無償授權使用，因此受到政府機構和非營利組織的高度評價；對機構而言，提供 LOD 是參與語意網環境的具體方式。在語意網技術框架中，基本應用主要是網路資料處理與查詢，進階應用還包括異質性的資料整合與知識推理等，為了結合實務現況與實驗語意網的效益，本研究利用「原住民文物典藏目錄」資料集為基礎，開發兩項應用包括：(i) 跨 Web 資料整合，這項應用將匯集網路上與原住民相關的政府開放資料，藉由 SPARQL 聯合查詢實現異質性整合。(ii) 語意網的知識推理，這項應用為解決資訊檢索可能發生的問題，例如關鍵字未知和知識概念不足等，將以建立知識本體（ontology）模型來描述知識定義，藉由語意推論器（reasoner）執行推理，以生成新的資料關係及依據這些新關係查找資料層的文物資料。兩項應用將利用網路程式進行資料重組混搭，最後透過 Web-based 介面呈現結果給使用者。簡言之，機構能提供合規的 LOD 是語意網能否成功的重要關鍵，本研究藉由一個應用範例，模擬外界（亦即非機構成員）如何使用機構內的資料，在目前共享開放的資訊環境下，最終受惠的將是包括機構在內的所有大眾。

In recent years, semantic web has grown from concept to realization stage. With the progress of semantic web technology and their corresponding software, practical applications require interoperable data provided by content owners. Because linked open data (LOD) provides free license for public use, it is highly praised by government and non-profit institutions. For the institution, providing LOD is a concrete way to participate in semantic web environment. In the framework of semantic web technology, Web data processing is basic usage. Advanced applications such as data integration across the Web and knowledge inference are the core value of semantic web technology. In order to experiment semantic web technology in the present scenario, this study utilizes the data set of "the cultural collection Catalog of Taiwanese indigenous peoples" to build two

developments: (i) Integration of heterogeneous data sources across the Web. The development integrates related data sets of indigenous peoples and is implemented by SPARQL federated query. (ii) Semantic knowledge reasoning. This development is used to address some difficulties that frequently troubled information retrieval, such as unknown keywords and the shortage of knowledge concepts. Knowledge reasoning is implemented by creating an ontology model to describe formal definition. A semantic reasoner is used to infer the model to generate new data relationships that is further used as a query criteria to find the information of cultural collections. The developments finally use a Web-based interface to present data mashup to users. Overall, this study illustrated that the use of existing LODs by outsiders (non-institutional members) can actually benefit to institutions. Consequently, semantic web technology is a concrete way to help data achieving the goal of visibility, searchability and discoverability. Content owners should positively provide LOD to attract more reuse by outsiders.

【關鍵字 Keyword】

語意網；目錄資料；鏈結開放資料；資源描述框架；知識推理
Semantic Web; Catalog; Linked open data; Resource description framework;
Knowledge inference

壹、緒論

語意網 (Semantic web) 是在 2001 年由 Tim Berners-Lee 提出的概念，目的為解決應用系統不能處理 World Wide Web (以下簡稱為 Web) 資料內容的問題，亦即為「電腦」建立友善的相互操作 (interoperable) 環境，以利系統能解讀、處理、推理資料。語意網技術聚焦在傳統 Web 資料缺少的資料識別、語意描述、關係定義等，因此語意網並非取代 Web 環境，而是補足現有架構缺少的功能 (Bizer et al., 2009)。由於 Web 是一個開放環境，參與者的屬性或角色非常多元，例如平台工具、資源提供、系統開發、與終端使用等，為確保操作順利運行，所有參與者必須遵循統一的協定、規範 (specification)，

例如全球資訊網聯盟 (World Wide Web Consortium, W3C) 即是制定規範的主要機構之一；因此所謂的語意網「技術」，也可以理解為 W3C 訂定的相關規範，「語意網技術堆疊圖¹」將這些技術規範以層次方式描繪了組成輪廓，包括基礎的 Unicode、URI (Uniform Resource Identifier) 等，中、高階的資源描述框架 (Resource Description Framework, RDF)、知識本體語言 (Web Ontology Language, OWL) (Panetto et al., 2016) 等；簡言之，語意網技術已從趨勢與概念，進入到具體可實踐的階段。

由於語意網應用是在開放且非專屬的架構下，因此依賴各角色之間的協同合作，主要包括：(i) 作業環境面，涉及標準組織訂定的技術規範與軟體商支援的工具，目前 W3C 規範與可用工具的質量已日趨完備；(ii) 應用開發面，包括內容提供及系統建置，前者泛指各機構須製作語意網技術可用的資料，也應儘可能授權大眾使用 (Bechhofer et al., 2013)，後者則是開發者應善用語意網技術的特性將效益極大化。近年來，各國興起的開放資料 (Open data, OD) 運動，強調大眾可無償使用資料且不受著作或專利權限制，這種開放風潮結合鏈結資料 (Linked data, LD)，即成為所謂的「鏈結開放資料」 (Linked open data, LOD) (Bizer et al., 2009; Zuiderwijk & Janssen, 2014)。許多領域已開始導入 LOD，特別是擁有大量數位內容的機構組織，包括國際組織 (如 World bank)、政府 (例如英國政府資料)、媒體 (如 BBC 新聞廣播)、社群 (如 DBpedia)、專業學科 (例如醫學的 BioPortal) (Hogan et al., 2012; Shadbolt et al., 2012)，已收錄數以千計的 LOD 資料集的「鏈結開放資料雲²」可獲得更詳細的資訊，這些無償及開放授權的資料已為知識融合帶來顯著的效益 (Buil-Aranda et al., 2013)。

近年來，圖書館、檔案館、博物館 (Libraries, Archives, Museums，以下簡稱為 LAM 社群) 也逐漸重視語意網的應用效益，許多研究也探討如知識本體 (Ontology) 及 LD、LOD 等技術對 LAM 社群的影響，例如有些研究分析這些新興技術的效益，認為可具體提升館藏資料的可見性、可查找性和可發現性 (Cole et al., 2017)；有些研究基於這些新興應用是技術導向，因此調查 LAM 社群在經營管理、館員態度、教育訓練等衍生問題 (Raza et al., 2019; Wahid et al., 2018)；有些研究

1 Semantic Web stack, 參考 <https://www.w3.org/2003/Talks/0624-BrusselsSW-IH/26.svgz>

2 The Linked Open Data Cloud, 參考 <https://lod-cloud.net/>

指出 LAM 社群的館藏是「公共資產」，應以提供 LOD 方式開放給大眾利用，吸引外界使用以創造資料的價值（Cole et al., 2017; Georgieva-Trifonova et al., 2019; Latif et al., 2015; Lampert & Southwick, 2013; Yoose & Perkins, 2013）；簡言之，LAM 社群普遍肯定語意網技術可帶來正面的幫助，尤其是認同語意網技術能提升館藏資料對外分享的效益。另一方面，LAM 社群因為館藏資料都很龐大，現有資料通常是藉由轉換而非重新編製為 LD 或 LOD；然而許多研究指出，受限於人員的專業能力不足（Khan & Bhatti, 2018）、傳統的機讀編目格式與 RDF 不完全相容（Cole et al., 2013）、經費資源侷限、管理層支持程度不一等因素，這些挑戰造成發展進度緩慢，不僅出現在開發中的國家，也同樣發生在先進國家中（Rasmussen Pennington & Cagnazzo, 2019; Warraich & Rorissa, 2020）。

國內 LAM 社群在語意網的開發也面臨上述挑戰，特別是具有「開放」的資料集或資料端點仍不普遍，後續應用也受到限制。因此，為彰顯「語意網時代」的理想應用，例如外界使用者（非機構成員）如何替機構開創資料效益，本研究嘗試利用「台灣原住民文物典藏目錄³」（以下簡稱「典藏目錄」）的 LOD 進行開發，包括跨 Web 的資料整合與知識推理等兩項應用，前者檢驗資料分享與再利用的方式，後者檢驗如何提升資料到知識層的應用。本研究內容安排如下：第二節為文獻與技術探討，簡要說明語意網技術、知識層技術、資料類型、LAM 社群的鏈結資料；第三節實驗跨 Web 的資料整合與混搭應用，利用 SPARQL 技術將分散的 Web 資料進行匯集，再以 PHP 混搭資訊呈現至 Web 介面；第四節實驗語意網的知識推理，將設計一個以主題概念組成的知識本體模型，透過語意推論器（reasoner）推理新的資料關係，再經由 SPARQL 查詢將資料層的事實呈現；最後，第五節總結本研究的重點與建議。

貳、文獻與技術探討

一、語意網技術

傳統 Web 以文檔為主體（例如網頁），識別文檔須利用網址

3 資料集在 Datahub，參考 <https://datahub.ckan.io/dataset/ascdc/resource/4a3b64ca-af5d-4979-8264-ca174612042c>

（URL）標示及定位，文檔之間以超連結（hyperlink）形成所謂的「文檔網絡」（Web of documents），文檔中的資料沒有自己的網址，因此資料不支援系統的相互操作（interoperability）；語意網是以資料為主體，個別資料具有獨立的可識別名稱，資料之間可以連結形成所謂的「資料網絡」（Web of data），因此使用資料就可以進行相互操作（Alvarez-Rodríguez et al., 2014; Hogan et al., 2012）。如前面所述，語意網技術是由規範組成，表 1 列示部分核心規範名稱及參考網址，以下簡述其意義與用途：

表 1
語意網部分核心規範

規範名稱	參考網址
URI（統一資源標識符）	https://www.w3.org/TR/uri-clarification/
RDF（資源描述框架）	https://www.w3.org/TR/rdf11-concepts/
SPARQL（RDF 查詢語言）	https://www.w3.org/TR/rdf-sparql-query/
OWL（知識本體語言）	https://www.w3.org/TR/owl2-overview/

- URI：用於標示 Web 資源名稱以提供識別，類似標示網頁的 URL，由於 URI 可以更完整仔細的標示資源，W3C 技術規範已全面使用 URI 代替 URL。
- RDF：此規範是一個資料結構（或模型），一個 RDF 陳述（statement）由主詞（subject）、述詞（predicate）、受詞（object）等三項單元組成，因此 RDF 陳述也稱為 RDF triple（以下稱為 RDF 三元組）。主詞須使用 URI 表達資源名稱以提供網路識別；述詞用於描述內容的屬性，通常是選用具有 URI 的 RDF 詞彙；受詞則是實際內容值，如果使用 URI 資源為內容值時即成為 LD，代表 RDF 陳述可以再連結其他資源（Marshall et al., 2012）。
- SPARQL：專為查詢 RDF 資料而訂定相關語法與函數的規範，作用類似 SQL 查詢關聯資料庫。SPARQL 使用 HTTP 協定傳送，因此相容於 Web 環境。SPARQL 1.1 新增聯合查詢（federated query）規範，查詢對象可遍及網路上任何 RDF 資料端點（endpoint）（Buil-Aranda et al., 2013）。
- OWL：用於建立知識本體模型塑模知識的規範，利用知識本體的概念化與階層原理表達知識組成，由於 OWL 與 RDF 語法相容，因此 OWL 的知識本體模型可結合 RDF-based 資料集進行推理。

二、語意網的知識層技術

為因應網路知識與智慧化發展，OWL 是語意網技術在「知識層」的主要規範，用於補足 RDF 在描述語意所欠缺的關係定義、規則運算、邏輯推理等（Motik, 2012; Zhou et al., 2016）。由於開發知識本體模型通常結合學理、OWL 規範、及軟體工具（如 Protégé、OntoEdit、TopBraid Composer 等），相近的術語經常混合使用，以下整理常用的知識本體構成元素，說明其意義及通用概念的術語：

1. 類別：具有相同特性事物的集合，通用概念的術語包括 *concept*、*category* 及 *class* 等，OWL 及軟體工具將「類別」稱為 *class*。
2. 屬性：用於描述類別特性，通用概念的術語如 *property*, *attribute* 等，當用於描述與其他類別的相關性時也使用關係（*relation*），OWL 及軟體工具將「屬性」稱為 *property*。
3. 實例：通常是指類別中的基本事物，具有類別定義的屬性，通用概念的術語如 *object*, *individual*, *instance* 等，OWL 及軟體工具將「實例」稱為 *individual*。

利用 OWL 編輯模型應依據知識本體原理進行⁴，例如分類（*classification*）是將相同的實例群用一個代表性的類別表示，因此使用類別「樂器」統稱各種不同的音樂器材，這種化繁為簡即是一種抽象化，有助於後續以邏輯關係來定義知識；知識本體另一個重要原理是「追本溯源」，它強調事物的形成都可以對應至階層化架構，由底層可以循序向上找到源頭，許多學科也普遍使用這種階層化並有不同的術語，例如在邏輯上稱為包容關係（*subsumption*）、資訊或知識領域稱為繼承或表示為“*is-a*”，知識內涵也可能是橫向的組合關係（*composition*）或表示為“*has-a*”。最後，OWL 為補充公理式（*axiom*）表達如命題、量化、集合等語意，因此規範也結合了描述邏輯（*description logic*）。

三、語意網的相關資料類型

隨著語意網應用普及，不同名稱的資料類型經常交互出現，例如 RDF 資料、OD、LD、LOD 等，由於資料類型對語意網技術的作業有不同意義，因此本研究針對它們的用途及差異予以釐清如下：

4 OWL 2 Web Ontology Language Primer，參考 <https://www.w3.org/TR/2012/REC-owl2-primer-20121211/>

- RDF 資料：指資料是 RDF 三元組（triple）格式，亦即依照主詞、述詞、受詞排列編輯，SPARQL 查詢必須使用 RDF 格式的資料（Breslin et al., 2010），非 RDF 資料可利用軟體轉換格式，W3C 已提供可利用的轉換工具清單⁵。
- OD：是一個概念性的指導方針，泛指任何已提供開放授權給大眾使用的數位內容，因此沒有限定資料格式（Attard et al., 2015）；例如國內政府資料平台⁶的資料集中，常見如 csv、json、excel、xml 等結構化格式，甚至也包含非結構的 pdf、word 等文書格式。
- LD：泛指符合 RDF 結構且資料內容包含與其他資料集之間的連結，它是實踐 Web of data 的具體方法，因此 W3C 鼓勵機構建立這類資料以避免產生資料孤島（data silos），造成應用受限（Bizer et al., 2009）。
- LOD：是指 LD 同時也是 OD，這類資料營造最友善的語意網使用環境，提供大眾自由應用與開發（Raza et al., 2019）。

除了上述資料類型的差異外，Tim Berners-Lee 也以「開放」為基本前提，提出利用資料的「五星等評級原則」，如表 2 所示，每星等涵蓋前面星等的內容，因此愈多顆星代表資料品質愈好，表中三星等評級（含）以下為一般 OD，四、五星等評級分別對應前述的 LD 及 LOD（Southwick, 2015）。簡言之，OD 是概念指導、RDF 資料是資料結構、LD 是實踐語意網的資料方法，而 LOD 則是集成上述所有特性的資料。

表 2
開放資料五星等評級原則（參考 <http://5stardata.info/>）

評級	意義說明
一星等	放置於 Web 上任何形式的電子資料，並以開放授權釋出。
二星等	提供機器可讀的特定結構化格式，例如 excel。
三星等	使用非專屬的通用格式，例如 csv。
四星等	採用 W3C 的標準（如 URI, RDF）來識別與編輯。
五星等	RDF 資料內容須建立與外部資料的鏈結。

5 RDF 轉換器 (ConverterToRDF)，參考 <https://www.w3.org/wiki/ConverterToRdf>

6 開放政府資料平台，參考 <https://data.gov.tw/>

在上述各資料類型中，LD 雖然是語意網技術的具體實踐，但外界在使用資料時仍有權限問題，LOD 因為結合了開放（Open）的精神，使得語意網能發揮最大的應用潛力（Yoose & Perkins, 2013）；因此，LAM 社群在導入語意網技術時多以開發 LOD 為目標，例如知名的 Smithsonian 圖書館（Thompson & Richard, 2013）、歐洲數位圖書館（Europeana），伊利諾大學圖書館（Cole et al., 2013）、內華達大學圖書館（Southwick, 2015）。所謂的 LOD，外界通常經由下列方式取得利用：(i) 公用的資料端點，例如 Europeana 透過 “<http://sparql.europeana.eu/>”，提供各界以 SPARQL 查詢 LOD；(ii) 下載具有開放授權的資料集，例如前述的「典藏目錄」已宣告開放授權（CC-0）。因此，對資料的使用者而言，LD 與 LOD 是非常不同的資料級別。

四、LAM 社群的鏈結資料

W3C 在 2011 年發布關於 LAM 社群的 “Library Linked Data Incubator Group Final Report”，用來幫助社群成員瞭解語意網技術及 LD 的關係，以利創建和管理有價值的館藏資產，該報告對 Library 的定義即是包含圖書館、檔案館、博物館，因此 Library Linked Data 的意義即為 LAM 社群內，任何以 LD 形式表達的館藏資產。由於 LAM 社群對資料有自己的傳統概念與術語，這些術語與 LD 有些不同但概念並不衝突，因此該報告主要是以 LAM 社群慣用的術語來對應 LD 的概念，另一份附錄報告⁷特別針對實踐上需要的資料集（Datasets）、詮釋資料元素集（Metadata element sets）、及資料值詞彙（Value vocabularies）提出說明，並列舉可用的資源作為 LD 建置使用，重點整理如下。

- 資料集：泛指 LAM 社群內各種形式的館藏紀錄集合、書目、目錄等，例如英國國家書目、公共圖書館的目錄、CrossRef 和 Europeana 等。
- 詮釋資料元素集：用於描述實體的類和屬性的詞彙，例如都柏林核心集（Dublin core），機讀格式（MARC）等，除了 LAM 社群的詮釋資料，語意網環境常用的詞彙集如 RDFS, OWL, SKOS, FOAF 等，也可以視為詮釋資料元素集的用途。

7 Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets, 參考 https://www.w3.org/2005/Incubator/llid/XGR-llid-vocabdataset-20111025/#Simple_Knowledge_Organization_System_.28SKOS.29

- 資料值詞彙：泛指索引典、主題詞、分類法、控制詞彙等，例如美國國會圖書館主題詞、農業控制詞彙（AGROVOC）、藝術與建築索引典（Art & Architecture Thesaurus, AAT）、Dewey 分類法、和地理名稱（GeoNames）。

本研究使用的「典藏目錄」可對應至上述的「資料集」，該資料集在述詞使用都柏林核心詮釋資料組織（Dublin core Metadata Initiative, DCMI⁸）的詞彙，因此述詞的角色即是對應至上述的「詮釋資料元素集」，最後該資料集在受詞使用 AAT 的詞彙如 *aat:300266639*，該受詞的角色是對應至上述的「資料值詞彙」。以下介紹 DCMI 及 AAT 的簡單背景與用途：

- 都柏林核心集（Dublin Core, DC）是由 DCMI 發展與維護的核心元素集，由於通過多個標準組織的認證（如 ISO, ANSI, IETF），許多非 LAM 領域也採用這些術語描述屬性。DCMI 在 2008 年成為獨立組織，持續發展標準化的詮釋資料，在發佈的詮釋資料標準中，DCMI Metadata Terms 是包含原始 15 個核心元素在內的擴充詞彙，每個術語都是 LOD 並利用「命名空間⁹」編排（Southwick, 2015; Yoose & Perkins, 2013），例如術語 Subject 的 LOD 表達 *http://purl.org/dc/terms/subject*，實務應用也常以命名空間（前置詞）縮寫為 *dc:subject* 或 *dct:subject*。
- AAT 是用於描述物質文化的控制詞彙，常用於編輯博物館的館藏目錄。AAT 在併入 Getty Trust 後也加入改造計畫，目前 AAT 擁有超過 5.5 萬個術語詞彙，並已全部轉換為 LOD 型式，術語之間的廣、狹義關係（broad-narrow）以上下階層表達（Cobb, 2015）。AAT 的命名空間為 *http://vocab.getty.edu/aat/*，個別術語使用編號表達，例如術語 clothing 為 *http://vocab.getty.edu/aat/300266639*，實務應用常以前置詞縮寫為 *aat:300266639*。查詢 AAT 術語有兩種方式如圖 1 所示，包括：圖左側為術語查詢介面¹⁰（以英文術語查詢為主，部分術語可使用中文）；圖右側為另一種查詢方式，使用者須在該網站提供的 SPARQL 端點¹¹，撰寫指令進行查詢。

8 Dublin core Metadata Initiative(DCMI)，參考 <https://dublincore.org/>

9 命名空間 (namespace) 是以 URI 描述的網路資源，為簡化 URI 的表達，命名空間通常以前置詞 (prefix) 縮寫。

10 術語查詢介面，參考 <https://www.getty.edu/research/tools/vocabularies/aat/>

11 Getty SPARQL 端點，參考 <http://vocab.getty.edu/sparql>

戚玉樑、丁嬋娟：語意網技術在館藏目錄中的應用：
以原住民文物典藏目錄資料為例

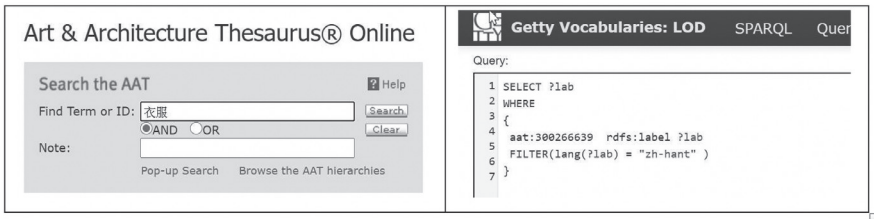


圖 1 AAT 檢索網頁與詞彙 SPARQL 端點

參、跨 Web 的資料整合與混搭應用

語意網技術為應用系統開發建立了可靠的環境，LOD 則創造共享的素材來源，本節以整合原住民相關資訊為目標，模擬收集不同性質的資料，包括組織機構、LAM 社群、鏈結開放資料雲等，圖 2 左側雲朵圖分別以我國政府資料、典藏目錄、DBpedia 代表這些異質來源，各資料以原住民族群或行政區做為關聯依據；圖 2 中間標示兩項作業：整合程序利用 SPARQL federated query 同時查詢不同端點；混搭程序利用 PHP 對查詢結果重新組合，最後以資訊混搭呈現給終端使用者。以下說明各資料的內容分析、前置處理及整合混搭的設計。

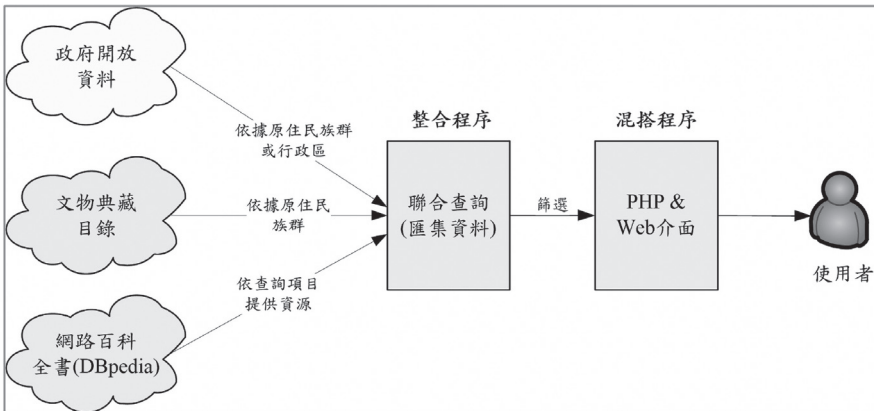


圖 2 語意網環境下的資料整合與混搭應用

一、資料源內容分析及前置處理

圖 2 左側標示的資料源皆為 OD，使用者可以進行後續重製、改作、編輯等利用，由於語意網技術要求資料須符合 RDF 結構，因此在

開發應用系統前，須分析資料內容及進行必要的前置處理，三項資料源分別說明如下：

1. 政府資料：國內政府資料開放平台目前已提供超過 4 萬個資料集，本研究由該平台下載與原住民相關的資料集，經分析後挑選包括人口統計、族群分布、原住民鄉鎮、文化機構等，由於資料集僅提供傳統格式，因此須進行前置處理包括轉換 RDF 格式、編輯述詞、編輯受詞為 URI 鏈結形式。
 - 轉換 RDF 格式：將政府資料集下載並匯入關聯資料庫，利用工具 D2RQ¹² 轉換為 RDF 資料。D2RQ 提供線上（online）及固定（persistent）兩種轉換方式，線上使用是指原始資料仍在資料庫，查詢是透過虛擬的 RDF 映射關係進行；固定使用是將資料匯出為 RDF 格式的資料集，提供其他軟體工具匯入。本研究採用固定方式並利用 Apache Jena Fuseki¹³ 進行資料集管理；上述類似功能的軟體工具很多，本研究選用它們的主因是普遍性及開放原始碼軟體（open source software）。
 - 編輯述詞：D2RQ 轉換資料庫時會同時產生初始的映射關係，其中述詞將直接套用資料庫中表的欄位名稱，但這類初始述詞通常不適合正規使用，開發者須編輯適當的述詞並重新執行關係映射。編輯述詞通常選用已存在且由權威機構發布的詞彙集（Myntti & Neatrou, 2016），例如 W3C 發布的 RDF, RDFS, OWL, SKOS, Vcard 等、機構組織提供的 Dublin core、Schema 及 DBpedia 等，開發者可利用線上工具查詢 RDF 詞彙，例如 LOV¹⁴ 是以單一詞彙查詢詞彙集、prefix.cc¹⁵ 則是以詞彙集名稱查找可用詞彙。
 - 編輯 URI 鏈結資源：原始資料的內容值通常都是文數字（literals），因此轉換至 RDF 的受詞也將是文數字。為提升資料的利用品質，應評估以 URI 資源取代文數字，例如以資源 <http://dbpedia.org/page/Taipei>（或 dbr:Taipei）取代文字型

12 D2RQ 是轉換關聯資料庫為 RDF 的開放軟體，下載參考 <http://d2rq.org/>。

13 Apache Jena Fuseki 是 Apache 基金會下的 Java-based 開放軟體，提供語意網和鏈結資料的開發

14 線上詞彙查詢 Linked Open Vocabularies repository (LOV)，參考 <https://lov.linkeddata.es/dataset/lov/>

15 線上詞彙查詢 Prefix.cc，參考 <http://prefix.cc/>

的「台北市」，這類規律性的更換也可利用編輯映射定義來批次處理，本研究已修改資料集中的行政區及原住民族群，將受詞內容值編輯為 URI 鏈結資源。

2. 典藏目錄：此資料是國內數位典藏計畫成果之一，資料集內容來自台灣數位檔案館聯合目錄，並已授權給大眾使用。此資料集擁有約 3,767 項典藏品目錄，經由轉換及編輯共建立約 10 萬筆 RDF 三元組的小型 LOD。附錄 1 是利用 SPARQL 指令查詢資料集的結果，本例摘錄一項典藏品目錄的 RDF 資料，內容呈現於一個主詞（<http://data.ascdc.tw/uc/anth/ntm/002f9ae3>）之下，包括同屬此資源的其他 RDF 三元組。附錄 1 的表格內已列示各 triple 的述詞與受詞，其中述詞部份以精簡的「命名空間（或前置詞）：詞彙」格式呈現，例如列號 8 的 *dc:subject*；受詞部分除一般的文字或數字外，須特別注意列號 1, 6, 8, 17, 25, 26 的受詞，它們是以 URI 表達的鏈結資源，例如列號 1 的 *dbr:Tsou_people* 可連結至 DBpedia 的「鄒族」，這些 URI 資源即是構成 LD 的關鍵。
3. DBpedia：內容是轉換自 Wikipedia 並以 LOD 呈現與編排，例如前述的 *dbr:Taipei* 與 *dbr:Tsou_people*，每項資源下由數量不等的 RDF 三元組組成，例如行政區的資源有簡介、各項地理及人文等資料，DBpedia 承襲 Wikipedia 具有多語言的特性，查詢時藉由語言代碼，例如英文（@en）、中文（@zh）等限定結果集。

二、資料整合與混搭應用

前述參與實驗的資料集，經內容分析及必要的前置處理後，已符合語意網技術對資料的相容要求，由於語意網將 Web 視為一個超大型資料庫，任何分佈於網路上的 LOD 資料集，皆可撰寫 SPARQL 查詢指令直接利用。

1. 資料整合：跨 Web 資料整合使用 SPARQL 聯合查詢，亦即在查詢指令加入”*SERVICE*” 函數發送至指定的端點執行。圖 3 是三項資料源整合的範例，本例是利用 Apache Jena Fuseki 軟體做為作業平台，兩個螢幕截圖分別包括上方的 SPARQL 查詢指令及下方的回傳結果。在 SPARQL 查詢指令部分，本例截圖省略前置詞定義，僅截取列號 16-33，其中列號 16 限定須依三項變數排列顯示，列號 18, 22, 28 分別向三個端點同時進行查詢。第一個端點內的查詢條件撰寫在列號 20，藉由特定主詞（代表一

項文物典藏品），查詢受詞（以變數 *?tribe* 表示），本條件的意義是查詢該典藏品所屬的原住民族群；第二個端點內的查詢條件包括在列號 24-25，將前面獲得的變數 *?tribe* 做為主詞，分別查詢 DBpedia 資源內的摘要（以變數 *?abstract* 表示）及資源標籤（以變數 *?label* 表示），列號 26 對內容分別限定取用中文及日文；第三個端點內，利用變數 *?tribe* 查詢該原住民族群分布的鄉鎮名稱（以變數 *?district* 表示）；最後，列號 33 限定結果為一筆。圖 3 下方顯示阿美族、鳳林鎮、及以日文的阿美族介紹，它們分別來自不同資料端點。

2. 混搭應用：圖 4 為利用 PHP 嵌入 SPARQL 查詢及執行資料混搭，再由網頁伺服器呈現，網頁截圖內容來自於政府資料集及 DBpedia，例如本例查詢行政區「花蓮縣」的原住民相關資訊，網頁左側由上至下分別包含：花蓮縣簡介、縣內原住民鄉鎮、縣內原住民部落等；網頁右側由上至下分別包含：花蓮縣總人口數與原住民人口數的比較、縣內不同原住民族群的人口分佈、縣內文化機構等。



圖 3 SPARQL 聯合查詢指令及結果

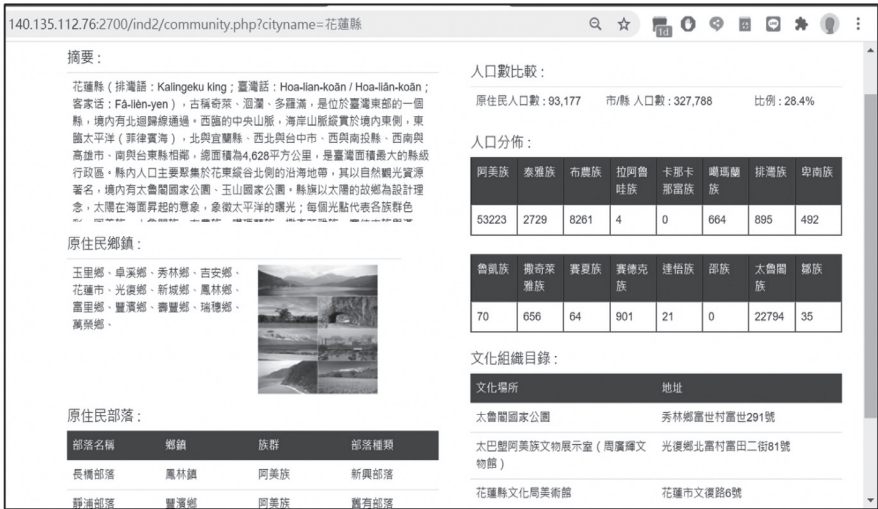


圖 4 利用 PHP 嵌入 SPARQL 查詢的資料混搭

肆、語意網的知識推理應用

前述第 3 節驗證語意網在處理 Web 資料與整合的效益，雖然 RDF 資料模型中的述詞已使用具有表達語意的詞彙，但描述範圍僅限於主詞與受詞之間的關係，亦即在同一個陳述內而不是資料之間，因此以 SPARQL 查詢典藏目錄資料集也面臨傳統資訊檢索的問題，包括：

1. 關鍵字未知：例如使用者不熟悉目錄資料的安排方式或因為時代不同也不習慣使用古典用語，例如木櫛、匏器、杵臼等，因此使用關鍵字查詢有相當難度。
2. 知識概念不足：例如資料集內容缺乏詞意邏輯或具有主題概念的定義，SPARQL 查詢也只能依賴字串比對進行查詢。

上述關於詞意邏輯的問題，可利用已提供 LOD 的索引典進行改善，例如 AAT 中的「樂器」是琴、笛、簫、弦、鼓等的上層詞彙，資料集如使用這些可供解析關係的 LOD 詞彙，SPARQL 即可直接使用具有廣義的詞彙進行查詢。本節以下針對解決主題概念的方式，設計如圖 5 所示的知識應用架構，主要是開發一個 OWL-based 的知識本體模型，利用邏輯式描述主題概念的定義，透過推論器推理新的資料關係，這些新關係將成為查詢條件並提供 SPARQL 對資料層進行查詢，最後整合結果並提供

給終端使用者。圖 5 的有向箭頭線僅表達資料流由左至右，亦即由資料層到使用者。然而在使用者的查詢應用中，執行請求（request）將循序由右至左執行。以下說明知識本體模型設計與應用系統開發。

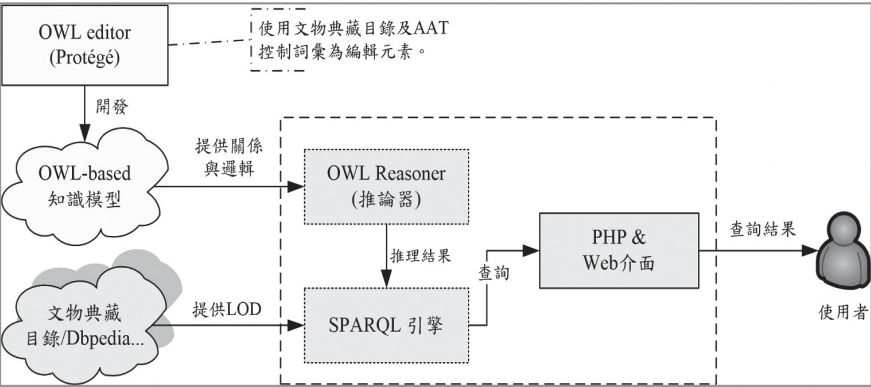


圖 5 語意網在知識推理的架構設計

一、知識本體模型設計

在資訊技術應用中，知識本體可概分為領域本體（domain ontology）及任務本體（task ontology）等類型，領域本體通常由大型權威機構訂定，提供廣泛及一般化的使用，因此須力求完整及周延正確；另一方面，任務本體是針對局部性問題而設計的解題模型，通常規模不大且配合應用系統解題，本研究以下開發的知識本體即採用「任務本體」型式。開發知識本體一般是利用軟體工具協助進行，目前有許多這類工具且通常提供整合開發環境，包括編輯器、偵錯、及各項檔案管理等，本研究選用 Protégé¹⁶ 進行知識本體開發，主要原因是擴充功能豐富多樣，例如可外掛（plugin）推論器及視覺化展示等。以下簡要說明利用 Protégé 建立知識本體的三項作業：

1. 建立知識本體雛型：知識本體由類別與階層化組成一個知識架構，利用 Protégé 建立時，區分為兩種類別，包括 (i) Primitive

16 Protégé 是由「史丹佛大學生物醫學資訊研究中心」研發的開放原始碼軟體，自 1999 年釋出至今已至 5.5 版，此工具允許安裝外掛套件擴充，因此在一些研究及網路上的評鑑中，均是最受歡迎的工具之一。

class，適用於設計基本類別，其定義的公理式應為必要條件；(ii) Defined class，適用於設計推理類別，其定義的公理式應為充分且必要條件（necessary and sufficient），這兩種類別在 OWL 語法分別對應至 *owl:Class* 與 *owl:equivalentClass*。基本類別是構成知識本體雛型的主要元素，圖 6 為 Protégé 建立類別的畫面截圖，左側視窗顯示本研究設計的典藏品類型及主題分類，其下包含由實心圓標示的基本類別，例如「典藏品_藝術品陶器雕刻」，主題分類下的基本類別是參考「台博館原住民文物分類表¹⁷」而命名，例如「工藝藝術」等 16 項。

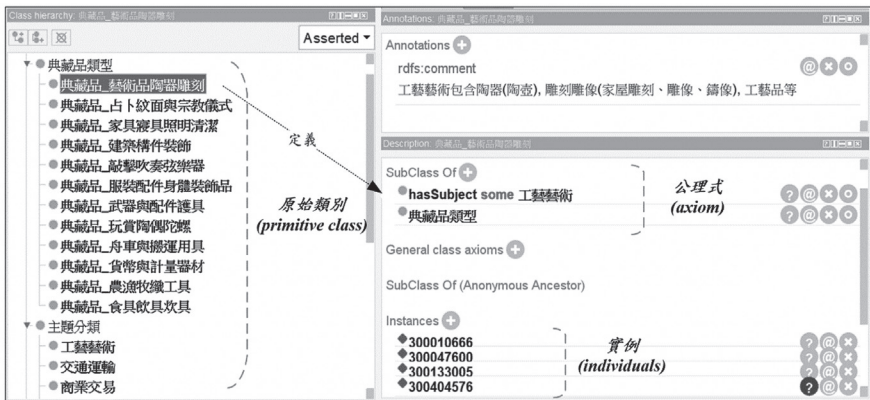


圖 6 Protégé 在編輯知識本體的畫面

2. 類別定義：有別於人類透過文字標籤理解類別意義，電腦系統的「理解」則是經由可解譯及運算的定義獲得。OWL 是知識本體的塑模語言，開發者須使用公理式編輯類別定義。前述基本類別定義僅使用必要條件，因此類別與公理式是單向包容關係（邏輯式為 $Q \rightarrow P$ ），例如圖 6「典藏品_藝術品陶器雕刻」的箭頭線已標示定義，規定了此類別下的實例必定具有公理式描述的條件，但反向則不成立。圖 7 左側視窗顯示六個推理類別，它們以含有全等符號（ $\equiv$ ）的實心圓表示，如「推理_精神生活」。本研究使用推理類別訂定所需的主題概念，例如精神生活主題泛指具有宗教、樂器、藝術、娛樂等撫慰心靈的用

17 國立台灣博物館人類學門原住民文物分類表，參考 <http://formosa.ntm.gov.tw/aborigine/class.pdf>

途，因此在「推理_精神生活」的定義中（箭頭線標示），已將四項基本類別進行聯集，由於推理類別的公理式需使用充分且必要條件，推理可以雙向成立（邏輯式為 $Q \leftrightarrow P$ ），因此本例特別重要的用途為推理四項基本類別下的任何實例，都將符合精神生活的文物。

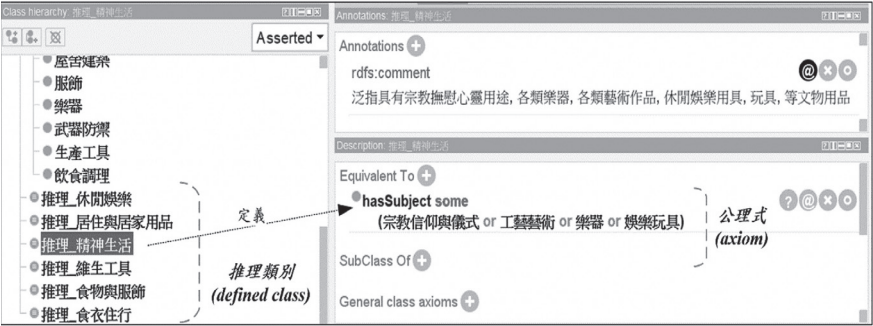


圖 7 推理類別使用充分且必要條件定義公理式

3. 實例宣告：為將知識層的知識本體模型與資料層的「典藏目錄」資料集產生聯繫，本研究以資料集使用的 AAT 術語做為實例並宣告至知識本體模型中，部分原資料集欠缺的術語已補充，以利實驗進行。在前述的圖 6 中，畫面右下是指定實例至「典藏品_藝術品陶器雕刻」基本類別中，例如包括 *aat:300010666* 在內的 4 項實例。實例因繼承類別的定義，當類別經推理而產生新關係時，實例也將自然繼承此關係。

完成上述知識本體模型後，利用 Protégé 附掛的推論器可進行驗證，圖 8 以橢圓虛線框標示為執行推理後（inferred）狀態，每項推理類別將依據公理式，運算符合條件的類別及其所屬的實例（亦即 AAT 術語），例如圖 8 左側以矩形虛線框標示新產生的類別架構，顯示推理後的「推理_精神生活」與圖 7 僅有單一類別明顯不同，包括新增「典藏品_占卜紋面與宗教儀式」、「典藏品_藝術品陶器雕刻」、「典藏品_敲擊吹奏弦樂器」、「典藏品_玩賞陶偶陀螺」等，後兩項也因符合「推理_休閒娛樂」的公理式條件，因此一併置放於此。矩形虛線框內的四項基本類別皆擁有實例，依據推理後的繼承關係，這些實例也都成為「推理_精神生活」類別的實例。

戚玉樑、丁嬋娟：語意網技術在館藏目錄中的應用：
以原住民文物典藏目錄資料為例

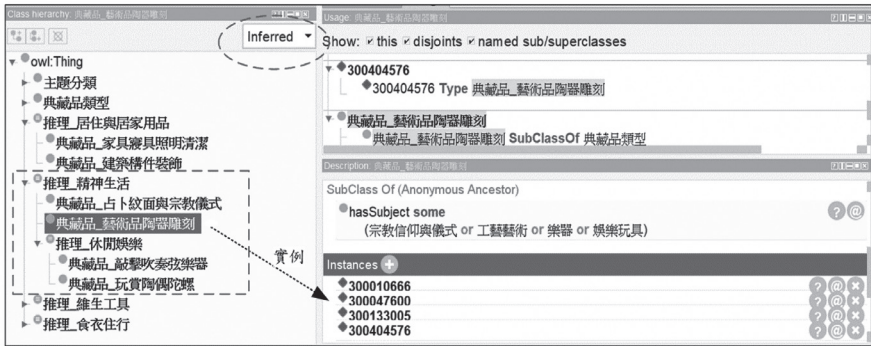


圖 8 知識本體模型的推理驗證

二、應用系統開發

圖 9 展示應用系統在知識推理的 Web 介面，為便於解說系統的設計，圖左側分別標示 (1) ~ (3)，其中標示 (1) 設計了兩欄下拉式選單，包括：「原住民族群」選單提供目前政府認定的 16 項族群名稱，例如最常見的阿美族到人數最少的卡那卡那富族；「主題分類」選單提供 18 個選項，其中 12 項為基本主題（如服飾、生產工具等）及 6 項具有推理功能的複合主題（如推理_精神生活、推理_食衣住行等），兩個選單內的選項共組為查詢條件，標示 (2) 與 (3) 為對應的執行結果，以下說明選單與後端系統設計的關係：

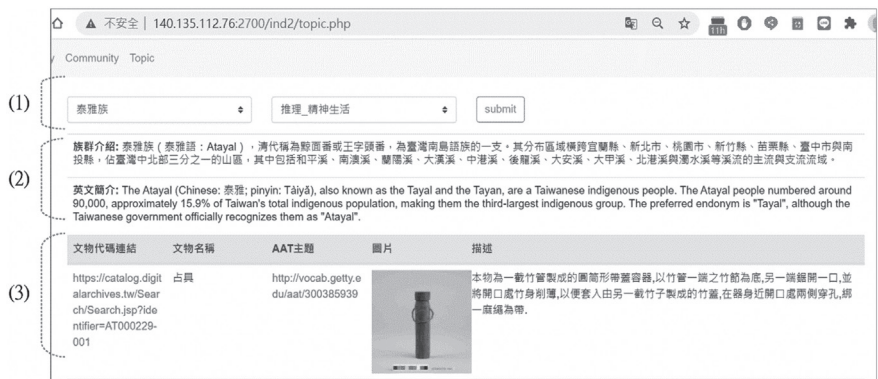


圖 9 利用 Web-based 應用系統展示資料混搭及知識推理

- 原住民族群：例如使用者選擇「泰雅族」，系統將傳遞 *dbp:Atayal_people* 至下列指令中的變數 *?ethnicity*，本指令是 SPARQL 聯合查詢，

將請求 DBpedia 回應 *?ethnicity* (亦即泰雅族) 的摘要及簡介，並分別限定中文與英文，查詢結果即為圖 9 標示 (2)。

```
SERVICE <http://dbpedia.org/sparql>
{
  ?ethnicity dbo:abstract ?abs;
               rdfs:comment ?comment.
  Filter (lang(?abs)="zh")
  Filter (lang(?comment)="en")
}
```

- 主題分類：例如使用者選擇「推理 _ 精神生活」，系統將傳遞 ind:推理 _ 精神生活至下列指令中的變數 *?topic*，配合前面選擇的「泰雅族」，此查詢的意義即為「泰雅族精神生活的文物有哪些？」下列指令使用 SPARQL 函數語法如 bind、concat、encode 等，它們可將資料合併、連結及編製 URI 格式。第 1 列指令由 SELECT 限定查詢結果集僅取出 5 類資料，圖 9 標示 (3) 分別依序對應如文物代碼連結 (*?DA_id*)、文物名稱 (*?lab*)、AAT 主題 (*?subject*)、圖片 (*?imgurl*)、描述 (*?desc*)；第 3 列指令“*?s rdf:type ?topic*”可觸發推論器，其中變數 *?s* 將獲得符合條件的實例 (AAT 術語)；第 4 列指令請求存放「典藏目錄」的端點執行聯合查詢；第 6 列指令“*?s1 dc:subject ?s*”將前述由推論器取得的變數 *?s* 傳遞至此並做為受詞，完成知識層與資料層的連結。

```
SELECT ?DA_id ?lab ?subject ?imgurl ?desc
WHERE {
  ?s rdf:type ?topic.
  SERVICE <http://140.135.112.76:3030/TDB2_0606/sparql>
  {
    ?s1 dc:subject ?s;
        ascdcanth:hasRelatedTribe ?ethnicity ;
        rdfs:label ?lab;
        dct:identifier ?id ;
        dc:subject ?subject;
        rdfs:comment ?desc.
    ?s2 edm:aggregatedCHO ?s1;
```

```
edm:isShownAt ?url.  
FILTER (isURI(?subject))  
FILTER (CONTAINS(?id, 'A'))  
Bind(URI(Concat(  
'https://catalog.digitalarchives.tw/Search/Search.jsp?  
    identifier='  
,ENCODE_FOR_URI(?id))) AS ?DA_id)  
}  
?s3 edm:isShownAt ?url;  
    rdf:type <http://example.org/vocab/imgurl>;  
    edm:hasView ?imgurl.  
}
```

圖 9 的展示畫面已完成圖 5 的設計，在推論器、SPARQL 引擎、Web Server 的協作下，已將知識本體模型與典藏目錄資料集結合以支援主題概念查詢，圖 9 僅選取文物目錄的部份資料，實務上可視情況調整查詢條件與顯示更多的目錄細節。

伍、結論、限制與建議

一、結論

本研究已利用語意網技術完成資料整合與知識推理等應用，使用者透過 Web-based 介面即可驅動相關檢索。(i) 在資料整合部分：圖 2 標示了包括政府機構、社群與典藏目錄等不同資料來源，在相容於語意網技術的資料格式下，利用整合程序、混搭程序，協助使用者獲得查詢資訊；圖 3 以樣板範例說明聯合查詢的用途，展示三項分屬網路不同端點的資料可以同時進行查詢；圖 4 是將聯合查詢結果混搭顯示，在選擇某一個行政區的查詢條件下，取得多項不同類型資料，包括行政區簡介、原鄉到文化組織等，顯示本系統能整合比其他類似網站更多元的資訊。(ii) 在知識推理部分：圖 5 標示了知識本體模型、推論器與資料層的運作關係，由於語意網的知識推理常利用 OWL 技術建立知識本體模型，因此第 4 節詳細介紹知識本體開發、知識本體模型推理及模型與資料合併應用；圖 6-7 說明利用 Protégé 開發知識本體模型的方式；圖 8 說明應用推論器進行知識本體模型的推理驗證，亦即檢

查預設的知識定義是否產生新的資料關係；圖 9 是將知識推理與代表資料層的「典藏目錄」一起運作，使用者可藉由概念方式檢索原住民文物資訊。最後，本研究也驗證語意網技術在資料共享與再利用的積極意義，那就是前述的系統開發可在沒有「內容所有者」的協助下，利用語意網技術與可用的資料即可在實驗室完成。

二、研究限制與建議

語意網應用須依賴各角色的協同合作，本研究藉由實作「台灣原住民文物典藏」的議題，模擬不同角色的分工與資料加值方式。有別於常規的系統開發通常有針對性或在局部範圍進行，本研究的系統開發主要是用於範例說明，模擬在 Web 的大範圍架構下，多個角色如何協同、進階應用如何關聯等，藉以強調資料品質對後續應用的影響；因此，本研究未採用常規的系統評估方式，而以完成上述各角色之間的協同合作取代，未來此範例擴充為正式系統時，可針對軟體品質 FURPS¹⁸ 如功能性、效能性等進行評估。最後，本研特別針對機構，例如政府部門或 LAM 社群等，提出下列建議：

1. 積極提供合格的資料：語意網需要 RDF-based 資料就如同傳統 Web 須要 HTML-based 內容網頁一樣，否則即使有完善的架構環境，沒有素材也會使應用成效不彰。國內可用的語意網資料確實相當缺乏，例如政府資料開放平台截至 2020 年底，僅有零星的資料集具有四星等評級，比率不到千分之一；因此，開發者在使用 OD 之前，必須重複對資料進行轉換作業，既不經濟也沒有效率。英國政府對 OD 的態度是一個模範，他們在 2007 年就推動政府釋出的資料應包括 RDF-based 格式在內，以利各界更有效率的利用（Shadbolt et al., 2012）。
2. 體認資料共享的意義：在過去年代中，因為資料壟斷或系統封閉，機構就是資源的擁有者與支配者，因此扮演包括編製、保管到系統開發等工作，應用範圍也受到侷限；然而在語意網環境下，機構內的資料可以藉由 LOD 分享給外界，吸引大眾利用也促使資料達到可見性、可查找性和可發現性的效用。機構須體認 LOD 的品質愈好愈能吸引大眾使用，甚至替機構開發

18 FURPS 是由五項屬性組成的軟體品質模型，包括功能 (Function)、可用性 (Usability)、可靠度 (Reliability)、效能 (Performance) 及可支援性 (Supportability) 等。

多元的應用，例如名為 See UK (<https://apps. seme4.com/see-uk>) 的應用系統，就是以英國政府的 LOD 開發如各地區犯罪類別、數字和比較等生活應用。簡言之，各機構要摒棄傳統資源支配者的思維，盡可能去除使用障礙以獲取外界的使用興趣，達到分工合作互利互惠的語意網理想。

誌謝

作者感謝陳昱任先生及蘇怡菁小姐協助應用系統開發，另外也感謝科技部提供經費支持（計畫編號 MOST 110-2410-H-033-032 -）

參考文獻

- Alvarez-Rodríguez, J. M., Labra-Gayo, J. E., & de Pablos, P. O. (2014). New trends on e-Procurement applying semantic technologies: Current status and future challenges. *Computers in Industry*, 65(5), 800-820. <https://doi.org/10.1016/j.compind.2014.04.005>
- Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, 32(4), 399-418. <https://doi.org/10.1016/j.giq.2015.07.006>
- Bechhofer, S., Buchan, I., De Roure, D., Missier, P., Ainsworth, J., Bhagat, J., Couch, P., Cruickshank, D., Delderfield, M., Dunlop, I., Gamble, M., Michaelides, D. Owen, S., Newman, D., Sufi, S., & Goble, C. (2013). Why linked data is not enough for scientists. *Future Generation Computer Systems*, 29(2), 599-611. <https://doi.org/10.1016/j.future.2011.08.004>
- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data: The story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1-22. <https://doi.org/10.4018/jswis.2009081901>
- Breslin, J. G., O'Sullivan, D., Passant, A., & Vasiliu, L. (2010). Semantic Web computing in industry. *Computers in Industry*, 61(8), 729-741. <https://doi.org/10.1016/j.compind.2010.05.002>
- Buil-Aranda, C., Arenas, M., Corcho, O., & Polleres, A. (2013). Federating queries in SPARQL 1.1: Syntax, semantics and evaluation. *Web Semantics: Science, Services and Agents on the World Wide Web*, 18(1), 1-17. <https://doi.org/10.1016/j.websem.2012.10.001>

- Cobb, J. (2015). The journey to linked open data: The Getty vocabularies. *Journal of Library Metadata*, 15(3-4), 142-156. <https://doi.org/10.1080/19386389.2015.1103081>
- Cole, T. W., Han, M. J., Weathers, W. F., & Joyner, E. (2013). Library MARC records into linked open data: Challenges and opportunities. *Journal of Library Metadata*, 13(2-3), 163-196. <https://doi.org/10.1080/19386389.2013.826074>
- Cole, T.W., Han, M. J. K., Janina Sarol, M., Biel, M., & Maus, D. (2017). Using linked open data to enhance the discoverability, functionality and impact of Emblematica Online. *Library Hi Tech*, 35(1), 159-178. <https://doi.org/10.1108/LHT-11-2016-0126>
- Georgieva-Trifonova, T., Zdravkov, K., & Valcheva, D. (2019). Application of semantic technologies in bibliographic databases: A literature review and classification. *The Electronic Library*, 38(1), 113-137. <https://doi.org/10.1108/EL-03-2019-0081>
- Hogan, A., Umbrich, J., Harth, A., Cyganiak, R., Polleres, A., & Decker, S. (2012). An empirical survey of linked data conformance. *Web Semantics: Science, Services and Agents on the World Wide Web*, 14, 14-44. <https://doi.org/10.1016/j.websem.2012.02.001>
- Khan, S, A., & Bhatti, R. (2018). Semantic Web and ontology-based applications for digital libraries: An investigation from LIS professionals in Pakistan. *The Electronic Library*, 36(5), 826-841. <https://doi.org/10.1108/EL-08-2017-0168>
- Lampert, C. K., & Southwick, S. B. (2013). Leading to linking: Introducing linked data to academic library digital collections. *Journal of Library Metadata*, 13(2-3), 230-253. <https://doi.org/10.1080/19386389.2013.826095>
- Latif, A., Scherp, A., & Tochtermann, K. (2015). LOD for library science: Benefits of applying linked open data in the digital library setting. *Künstl Intell*, 30(2), 149-157. <https://doi.org/10.1007/s13218-015-0420-x>
- Marshall, M. S., Boyce, R., Deus, H. F., Zhao, J., Willighagen, E. L., Samwald, M., Pichler, E., Hajagos, J., Prud'hommeaux, E., & Stephens, S. (2012). Emerging practices for mapping and linking life sciences data using RDF: A case series. *Web Semantics: Science, Services and*

- Agents on the World Wide Web*, 14, 2-13. <https://doi.org/10.1016/j.websem.2012.02.003>
- Myntti, J., & Neatrou, A. (2016). Use existing data first: Reconcile metadata before creating new controlled vocabularies. *Journal of Library Metadata*, 15(3-4), 191-207. <https://doi.org/10.1080/19386389.2015.1099989>
- Motik, B. (2012). Representing and querying validity time in RDF and OWL: A logic-based approach. *Web Semantics: Science, Services and Agents on the World Wide Web*, 12-13, 3-21. <https://doi.org/10.1016/j.websem.2011.11.004>
- Panetto, H., Zdravkovic, M., Jardim-Goncalves, R., Romero, D., Cecil, J., & Mezgar, I. (2016). New perspectives for the future interoperable enterprise systems. *Computers in Industry*, 79, 47-63. <https://doi.org/10.1016/j.compind.2015.08.001>
- Rasmussen Pennington, D., & Cagnazzo, L. (2019). Connecting the silos: Implementations and perceptions of linked data across European libraries. *Journal of Documentation*, 75(3), 643-666. <https://doi.org/10.1108/JD-07-2018-0117>
- Raza, Z., Mahmood, K., & Warraich, N. F. (2019). Application of linked data technologies in digital libraries: a review of literature. *Library Hi Tech News*, 36(3), 9-12. <https://doi.org/10.1108/lhtn-10-2018-0067>
- Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., Hall, W., & schraefel, m. c. (2012). Linked open government data: Lessons from Data.gov.uk. *IEEE Intelligent Systems*, 27(3), 16-24. <https://doi.org/10.1109/MIS.2012.23>
- Southwick, S. B. (2015). A guide for transforming digital collections metadata into linked data using open source technologies. *Journal of Library Metadata*, 15(1), 1-35. <https://doi.org/10.1080/19386389.2015.1007009>
- Thompson, K., & Richard, J. (2013). Moving our data to the semantic web: Leveraging a content management system to create the linked open library. *Journal of Library Metadata*, 13(2-3), 290-309. <https://doi.org/10.1080/19386389.2013.828551>
- Wahid, N., Warraich, N. F. & Tahira, M. (2018). Mapping the cataloguing practices in information environment: A review of linked data

- challenges. *Information and Learning Science*, 119(9/10), 586-596. <https://doi.org/10.1108/ILS-10-2017-0106>
- Warraich, N. F., & Rorissa, A. (2020). Application of linked data technologies in libraries: Pakistani information professionals' attitudes and perceptions, *The Electronic Library*, 38(5/6), 1035-1051. <https://doi.org/10.1108/EL-01-2020-0002>
- Yoose, B., & Perkins, J. (2013). The linked open data Landscape in Libraries and Beyond. *Journal of Library Metadata*, 13(2-3), 197-211. <https://doi.org/10.1080/19386389.2013.826075>
- Zhou, Q., Simmhan, Y., & Prasanna, V. (2016). Knowledge-infused and consistent complex event processing over real-time and persistent streams. *Future Generation Computer Systems*. <https://doi.org/10.1016/j.future.2016.10.030>
- Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, 31(1), 17-29. <https://doi.org/10.1016/j.giq.2013.04.003>

附錄 1：文物典藏目錄資料集範例

	述詞 (predicate)	受詞 (object)
1	ascdcp:anth/hasRelatedTribe	dbr:Tsou_people
2	ascdcp:length	"2510.0"^^xsd:float
3	ascdcp:otherMeasurementInfo	"1 個"@zh
4	ascdcp:type	ascdct:Texts
5	dbo:currentStatus	"良好"@zh
6	dc:rights	<http://viaf.org/viaf/140029379>
7	dc:source	"國立臺灣博物館原住民文物數位典藏計畫網站"@zh
8	dc:subject	aar:300266639
9	dc:subject	"身體裝飾"@zh
10	dc:dataset	"國立臺灣博物館原住民文物典藏數位化計畫"@zh
11	dc:dataset	"聯合目錄"@zh
12	dcterms:description	"鄒族頸飾分為兩種，一為頸帶；另一為頸鍊係以貝珠、玻璃珠、銀鍊、銅鍊或豬牙等穿綴而成。兩種頸飾可同時穿戴 (李莎莉, 1998 : 149)。"@zh
13	dcterms:description	"由黑色孔珠串成。"@zh
14	dcterms:description	"繫於頸部，具有裝飾效果。"@zh
15	dcterms:identifier	"AT1513"
16	dcterms:identifier	"1513.00"
17	gvp:aar2218_used-function_as	aar:300209284
18	gvp:aar2427_produced_by	"串連"@zh
19	ov:shape	"環狀"@zh
20	rdf:type	ascdct:AscdcProvidedCHO
21	rdfs:label	"Necklace"@en
22	rdfs:label	"頸飾（玻璃珠頸飾）"@zh

	述詞 (predicate)	受詞 (object)
23	sam:samplingMethod	" 接收 "@zh
24	schema:color	" 黑 "@zh
25	schema:isRelatedTo	< http://collection.britishmuseum.org/id/object/EAS22425 >
26	schema:material	aat:300010797

The Application of Semantic Web Technology in Collection Catalog: Take the Cultural Collection Catalog of Indigenous People as an Example

Yu-Liang Chi

Professor

Department of Information Management
Chung Yuan Christian University

Chan-Chuan Ting

Instructor, Department of Accounting,
Chung Yuan Christian University

Ph.D. Student, Ph.D. Program in Business,
Chung Yuan Christian University

Introduction

The Semantic Web is to establish a user-friendly operating context for application systems, and to provide data identity, semantic description, and relation definition, which are lacking in current web technologies, for solving problems of data processing. Therefore, Semantic Web is not to replace the Web, but to make up for the deficiencies in the present framework. In recent years, Semantic Web has developed from a concept into an application. Institutions that store a large quantity of data, such as libraries, archives, museums (LAM), will benefit considerably from such an application. It enables the data to be visible, searchable and retrievable. The framework of Semantic Web requires the active involvement of various agencies, especially the institutions owning the data have to undertake more responsibilities and tasks than before. Therefore, when introducing Semantic Web into an institution, it is crucial to make aware of the fact that data users will not be limited to the members in that particular institution

or other homogeneous groups, but to include all users on the Web. With this awareness, one of the major tasks for the institutions is to provide data that conforms to the standards required by Semantic Web. It is observed in this study that there is a considerable gap between an ideal context for Semantic Web and the attitude of the LAM society toward data conversion, compilation and open authorization. The effectiveness of its subsequent applications thus has not been revealed yet. Therefore in this study an application model is developed to investigate how the LAM data in a Semantic Web context can be integrated with other data for users to search and use for advanced knowledge inference. Through a demonstration of using various applications, it is expected to emphasize the fact that data quality is the key to the effectiveness of Semantic Web.

Methodology

The Search and Knowledge Inference of Information Related to Indigenous People is set as the subject of the model development in this study. In addition to incorporating various data sources with Semantic Web technology, the Cultural Collection Catalog of Indigenous People issued by a museum is also adopted for developing knowledge inference. The collection is linked and open data conforming to Semantic web standards and has been authorized to be used for free. The two major applications include,

1. Integration and mashup of data across the Web

The model shown in Figure 1 simulates how to integrate various heterogeneous data in a Semantic Web context. Three major practices include selecting data sources, handling derivative problems, and developing practical applications. First, as shown in Figure 1, the data sources include open government data, cultural collection catalogs, and linked open data (such as DBpedia). Second, the major derivative problem is that partial data sources do not conform to Semantic Web standards, and thus have to be processed in advance, such as being converted into a standard format or being edited in fields of descriptions or linked sources. Lastly, the practical applications shown in Figure 1 are categorized into two procedures, semantic integration and semantic mashup. Semantic integration refers to programming the search function using SPARQL

language, especially the cross-web data integration using the federated query function. Semantic mashup refers to writing web pages using the PHP programming language, then mixing and displaying search results for end users to browse.

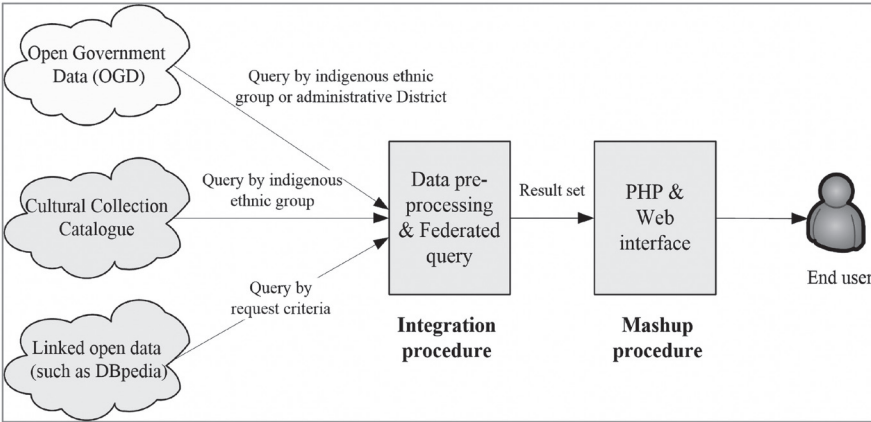


Figure 1 Data Integration and Mashup in a Semantic Web Context

2. Applications of Knowledge Inference

The model shown in Figure 2 demonstrates how to design an independent knowledge model, how to incorporate this knowledge model and data collections, and how to conduct knowledge inference using reasoners and SPARQL program language. First, the OWL editing software (such as Protégé) is adopted to develop an ontology model, with a Resource Description Framework (RDF), inferring new relations among data using reasoners. These new relations will become the query criteria for searching in the linked data of Cultural Collection Catalog of Indigenous People, which is programmed with SPARQL query language. The results are integrated with the PHP programming language, and provided to end users. The arrow line shown in Figure 2 indicates the data proceeding from left to right, that is, from the data layer to users; however, in applications of user query, the requests proceed from right (users) to left (the database).

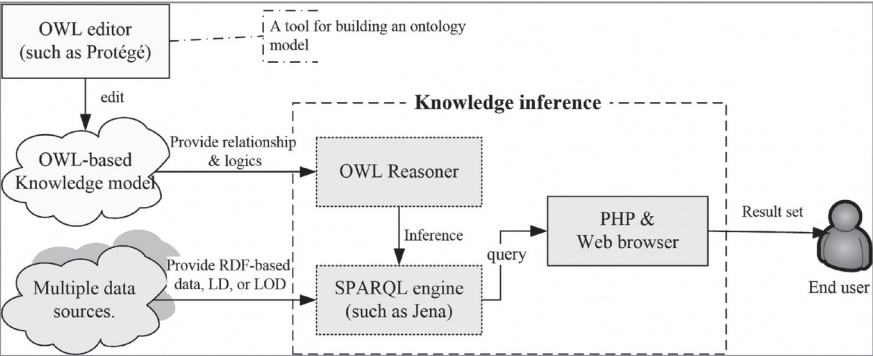


Figure 2. The Framework Design of Knowledge Inference in Semantic Web

Conclusions and Suggestions

In this study, a practical model is developed to simulate how external users (that is, non-institutional members) use the internal data of various institutions. This model is developed independently in the laboratory without assistance from other institutions, for according with the important principle of Semantic Web technology, that is, data sharing and re-using. In the past, institutions are owners and controllers of resources, thus also undertake the tasks of data compilation and storage, and system development and management, which might result in problems of resource monopoly or closed systems. However, with the influences of worldwide open data movements, institutions with data that is classified as public assets, such as governmental institutions or the LAM societies, have to share information for the public to use, and should also devote to enhancing data quality for conforming to the data standards of Semantic Web. If the data is intriguing enough to attract the public to use, the benefited parties will be the whole society including the institutions owning the data.