

鏈結資料於數位典藏之研究：以畫家陳澄波為例

陳淑君

中央研究院歷史語言研究所助研究員

E-mail: sophy@sinica.edu.tw

關鍵詞：陳澄波；鏈結資料；鏈結開放資料；開放文化資料；語意數位圖書館

【摘要】

鏈結資料促使文化資產機構的藏品與研究材料，可與全球的資料網路進行具有語意關係的連結與整合。本文以數位典藏資源為對象，選擇「從北緯 23.5°出發：陳澄波」主題網站的資料作為研究案例，進行鏈結資料的研究實作。本研究聚焦於人物類型的數位典藏資料，藉以展示人物及其相關典藏文物為核心的人、時、事、地、物之間的脈絡關係，並能夠基於語意關係回答複雜問題，提供視覺化呈現，讓大量的數位人文資料可以更便於由機器理解與處理，進而使研究資料能夠為使用者取用、再利用與重新組合，以提升數位化資料使用的效能。

前言

鏈結資料的概念是由全球資訊網 (WWW) 創始人 Tim Berners-Lee 於 2006 年首次提出，他當時著眼於國際間開放政府資料 (open government data) 的興起，提出可以在此基礎之上，讓原始資料 (raw data) 以鏈結資料的方法，加以存放與再利用 (Berners-Lee, 2009)。隨後，Tim Berners-Lee 以研究實例示範開放政府資料的再利用如何加惠於公眾，例如利用英國交通部的開放資料，並混搭 (mashup) 其他資料來源，諸如地圖，最後以互動式地圖呈現出最常發生腳踏車交通意外事故的地點 (Berners-Lee, 2010)。在開放政府資料的倡導計畫方面，以英國為例，其於 2009-2016 年間已經產生超過四萬套可供再利用資料集 (datasets)，主題包含環境、政府支出、犯罪與司法、教育、交通、地理等多種資料類型。相關文獻也顯示開放政府資料已影響當今社會的諸多層面，並引發大量的討論與研究 (Zuiderwijk, Janssen, van den Braak, & Charalabidis, 2012; Ubaldi, 2013; Jetzek, Avital, & Bjørn-Andersen, 2014)。鏈結資料在文化資產社群的發展，從 2011 年 6 月「圖書館、檔案館與博物館的鏈結開放資料」(Linked Open Data in Libraries Archives and Museums, 以下簡稱 LODLAM) 高峰會議在美國舊金山舉行第一次會議後，開放文化資產資料 (open cultural heritage data) 正式於國際間聚集並形成

具體的研究社群 (Voss, 2012)。LODLAM 於 2013 (加拿大蒙特婁)、2015 (澳洲雪梨), 2017 (義大利威尼斯) 年等, 持續舉行每二年一次國際高峰會議, 許多的研究個案與成果也在此會議公開。

回顧臺灣過去十五年 (2002-2016) 以來, 經由科技部數位典藏與數位學習國家型計畫、文化部國家文化資料庫, 以及後續相關永續維運計畫, 已經累積五百萬餘件數位典藏資源, 主要來源包括政府部會、文化資產機構及學術社群, 產出的數位資源主要包括國家型或機構型聯合目錄、主題資料庫及主題網站等不同類型系統。換言之, 這些數位材料涵蓋政府資料與文化資料, 也正是前文提及的開放資料運動主要對象。由於數位典藏的建置橫跨至少十五年, 歷經資訊科技典範不斷轉移與推進, 從 Web 1.0、Web 2.0、Web 3.0 甚至即將進入 Web 4.0。以 Tim Berners-Lee 提出的鏈結資料概念與典範, 可以觀察到過去已建立的典藏資源, 仍是傳統的文件網 (web of documents) 而非具備語意的資料網 (web of data) 之典範, 前者以文件為主、便於人類閱讀, 後者則是以資料為主、提供機器可讀取與處理的方法 (Bizer, Heath, Idehen & Berners-Lee, 2008)。

以數位典藏國家型計畫的人物及其相關文物的主題網站為例, 包括李澤藩素描創作手稿數位美術館計畫[1]、臺灣本土音樂家之影音典藏—李泰祥大師[2]、臺灣典範書家陳丁奇數位美術館[3]、打造福爾摩沙之聲: 張連昌薩克斯風紀念館數位典藏計畫[4]、史惟亮教授音樂數位典藏計畫[5]、臺灣民間藝術家數位藝術博物館之建置—以楊英風數位藝術博物館為例等[6], 都是以視覺藝術與表演藝術為主的數位典藏計畫。經逐一進入數位典藏的主題網站檢視後, 可以觀察到以人物及其相關文物的典藏, 通常包括生年表、相關作品、相關文物與蒐藏、相關文獻等內容, 但是這些資料身處在語意網環境卻面臨許多的局限, 例如資料在封閉式系統, 無法讓使用者依照感興趣的特定材料輕易地於全球資訊網查詢; 此外由於這些資料都是採取無法再延伸的字串 (string), 而非以能夠再延伸的事物 (thing) 為設計基礎, 因此不容易探索相關脈絡之間的資料。例如, 在李澤藩素描創作手稿數位美術館網站的生年表資料, 其中一項「1928 年, 李澤藩的作品〈夏日的午後〉入選台展; 同年, 廖繼春〈芭蕉之庭〉、陳植棋〈臺灣風景〉入選帝展」, 我們只能閱讀這些訊息, 但無法從系統進一步探索廖繼春的生平資料、其作品〈芭蕉之庭〉的相關資料或者帝展是何種性質的展覽會。

有鑑於此, 本文希望探索數位典藏的主題網站如何以鏈結資料的方法, 轉換為具有語意功能的數位系統。本研究聚焦於人物類型的數位典藏資料, 主要的研究目的是建構以鏈結開放資料為基礎的系統, 藉以展示人物及其相關文物藏品為核心的人、時、事、地、物之間的脈絡關係, 並能夠基於語意關係回答複雜問題, 提供視覺化呈現, 讓大量的數位人文資料, 可以更方便由機器理解與處理, 進而使研究資料能夠取用與再利用, 提供使用者得以重新組合, 進而提升數位化資料的效能。主要的研究問題包括:

- 一、鏈結資料如何強化數位典藏資源的探索性？
- 二、鏈結資料應用在數位典藏資源，可能提供的脈絡化方式？
- 三、鏈結資料如何自動推理產生具語意關係的資料？

相關研究

總體而論，圖書館、檔案館與博物館等文化資產社群在鏈結資料(Linked Data，或稱 Linked Open Data, LOD，本文將以 LOD 簡稱)方面的研究發展，包括資料模型、屬性語彙、屬性值語彙及藏品 LOD 等四大層面：

- 一、資料模型：意指資料如何被表達、儲存及取用的方式，包括資料的格式、定義和屬性，資料之間的關係，以及資料的限制等整體架構概念(林信成，2012)，在本研究中又經常被稱為具有特定領域的知識本體架構。例如美國國會圖書館的書目架構(BIBFRAME)(Kroeger, 2013; Library of Congress, 2016)，歐盟數位圖書館的資料模型(EDM)，以及國際博物館協會(ICOM)屬下國際文獻工作委員會(CIDOC)發展的概念參考模式(CIDOC CRM)等。
- 二、屬性語彙：意指可用以連結類別間或主體與客體間的關係屬性之詞彙，可以對應到傳統詮釋資料的元素或欄位，例如題名、創作者等資源的屬性。常見的標準包括 Schema.org、SKOS、DCMI Metadata Terms、FOAF(Friend-of-a-Friend)等語彙集。
- 三、屬性值語彙：意指類別、或主體與客體本身的資料值，且它們多具有可再延伸的屬性，例如某資源「蒙娜麗莎的微笑」具有「創作者」屬性，其屬性值是「達文西」，「達文西」語彙本身具有「生卒日期」、「國籍」、「作品」等屬性。此類型語彙多數來自文化資產機構的各種類型權威檔，如：美國國會圖書館的 LCSH、LCC、MARC Relators 等，以及美國蓋堤研究所(Getty Research Institute)的藝術與建築索引典(Art & Architecture Thesaurus, AAT)、藝術家人名權威檔(Getty Union List of Artist Names, ULAN)、地名索引典(Getty Thesaurus of Geographic Names, TGN)等控制詞彙。
- 四、藏品 LOD：意指以館藏目錄或資料為對象，採用上述的各式標準進行語意轉換，並以三元組的敘述(statement)方式取代傳統的書目或詮釋資料記錄(record)方式，提供資料再利用。如法國國家圖書館、大英圖書館、大英博物館、美國數位公共圖書館(DPLA)等，將館藏目錄轉換成具有語意的鏈結資料。

除了上述大型圖書館與博物館積極將其館藏以 LOD 方法進行語意轉換與發佈，成為全球資訊網的鏈結開放資料雲圖(LOD cloud diagram)的一部分(Abele, McCrae, Buiteiaar, Jentzsch, & Cyganiak, 2017)，以便各方使用者可以利用機器取其所需資料集並重新再利用之

外，許多文化資產機構與學術機構也以小型的 LOD 研究計畫進行實驗與探索，主要特色可以歸納如下：

以主題為例，展現多重來源資料的整合

此類特色是 LOD 計畫以一個特定主題為焦點，藉由消費其他不同來源的開放資料集，藉以展示語意整合的效益。案例包括：英國 OpenART 以藝術銷售為主題、美國 Linked Jazz 以爵士樂家為主題、義大利 Open Memory Project 以在義大利的猶太人大屠殺歷史為主題，以及美國 OCRE 以古羅馬錢幣為主題等計畫。以 OCRE 為例，該計畫全名是 Online Coins of the Roman Empire (OCRE) [7]，主要目標為鑑定、編目與研究羅馬帝國錢幣，由美國錢幣協會 (American Numismatic Society) 及紐約大學古代世界研究所 (Institute for the Study of the Ancient World) 共同開發，涵蓋所有已知羅馬硬幣類型的線上目錄。該計畫使用 LOD 技術，再將錢幣概念分別連結到擁有實體藏品的美國錢幣學會典藏 (ANS collection)、柏林錢幣博物館 (Münzkabinett of the State Museums of Berlin)、維吉尼亞大學藝術博物館 (University of Virginia Art Museum)、以及大英博物館 (British Museum) 等不同來源之目錄記錄。

圖書館、檔案館、博物館社群各自展現不同模式

雖然同為文化資產社群，但是圖書館、檔案館及博物館機構與資料在 LOD 的研究實驗仍然各自展現出不同的特質與模式。圖書館社群的目標相當一致，多是期待藉由 LOD 方法，將儲存於封閉系統，無法在全球資訊網被檢索到的資料，例如以 MARC 為基礎的圖書館目錄加以解放並轉換成具機器可處理的語意標準，進而爭取更多的讀者造訪與使用圖書館資源。檔案館社群雖然未如圖書館般，發展出本身的資料模型及標準，但是檔案資料卻被許多的 LOD 研究計畫所運用，並充份展現各種資料之間的脈絡關係。博物館社群基於文物資訊之間的互通有無之研究需求，但基於館際之間並不若圖書館長期以來，已經建立一致性的書目格式，因此 LOD 計畫強調採用具互通性的資料模型，藉以整合不同來源、類型與格式的資料。以下分別就此三個社群的 LOD 研究發展，分別進一步說明。

圖書館社群進行 LOD 計畫的主要目的，多是為了增進館藏目錄在全球資訊網的能見度與推廣。早期，實驗的主要目標是將館藏目錄轉為資源描述架構 (Resource Description Framework, 簡稱 RDF) 格式，例如法國國家圖書館進行 LOD 計畫的主要動機，是希望藏品資源傳播範圍更廣，並藉由與外部資源的鏈結，增加本身館藏在網路世界的再利用價值及能見度。該圖書館採用語意網技術及 LOD 原理，包括運用 RDF、提供 URLs、採用 FRBR 模型，以及 SKOS 等語彙標準，並將該館不同格式的各種目錄，諸如線上公共目錄 (INTERMARC 格式)、檔案資料 (XML-EAD 格式)，以及數位圖書館 (Dublin Core 格式) 等藉此加以整合 (Illien, 2012)。近期，由美國國會圖書館於 2011 年開始採用語意網和 LOD 方法，發展書目架構 (BIBFRAME) 計畫，嘗試探索未來的書目描述如何與網路世界的資源連結，並於 2012 年完成書目架構草案，

2016 年 4 月公布 BIBFRAME 2.0 版本後，以美國本土為主的圖書館開始展開先導實驗，以便未來能將機讀編目格式 MARC 紀錄轉至 BIBFRAME。例如：LD4P(The Linked Data for Libraries: Linked Data for Production) 計畫 (Durante, Weimer, & McGee, 2016)，主要成員包括：康乃爾大學、哈佛大學及史丹佛大學等圖書館。上述的資料模型與知識本體多是重量型 (heavyweight)，諸如：FRB 或 BIBFRAME，不過也有圖書館嘗試採用輕量型 (lightweight) 知識本體作為 LOD 計畫的標準語彙，例如：瑞典國家圖書館採用 FOAF、SKOS、BIBO、Dublin Core 等標準詞彙揭露資料，於 2008 年發佈為鏈結開放資料，是全球圖書館社群最早將書目發佈為鏈結資料的個案 (Byrne & Goddard, 2010)。上述提及的重量型或輕量型知識本體，主要差異在於資料模型的完整程度，以及類別與屬性的細緻程度。重量型知識本體嘗試盡可能地保留資料所有的脈絡描述，而輕量型知識本體則以最通用的類別與屬性進行資料描述。

檔案類資源則是多以人物或團體為主題建立 LOD 計畫，例如：Linked Jazz 及 Open Memory Project 等計畫分別曾獲得 2013 及 2015 年 LODLAM 高峰會議的 LOD 競賽作品首獎。Linked Jazz 計畫是由紐約普瑞特藝術學院圖書資訊學系組成的團隊，主要目標為提高數位文化資產資料的使用率及可見度，以及推廣 LOD 技術的應用與方法，該計畫運用爵士樂數位化歷史檔案資料，開發視覺化工具以呈現音樂家之間豐富的社交網絡關係，並以群眾外包 (crowdsourcing) 工具，邀請專家從取自歷史與訪談的檔案材料，選擇與辨識爵士藝術家之間的關係，實現檔案和數位資料間有意義的關聯 (Pattueli, 2012)。義大利 Open Memory Project 開放回憶計畫則是以二次大戰時期的猶太人大屠殺為主題，由義大利保存研究猶太人文史遺產的基金會 Fondazione Centro Di Documentazione Ebraica Contemporanea (CDEC) 應用 LOD 技術，將在義大利大屠殺中的猶太人歷史資訊，包含人物、經歷、事件，從分散在不同資料庫中的歷史檔案、照片、影音檔等，整合在一個資料庫的管理平台以供使用者檢索。該計畫從人名與傳記資料作為切入點分析資料，取得超過 19,000 條的人物記錄，30 多個關於每位受害者及其親屬的生平和遭受迫害資料的屬性，包括：姓名、生卒年、國籍、性別、父母姓名、宗教信仰、職業、逮捕地點、驅逐出境前的集中地點、到達納粹集中營紀錄、命運 (死亡、死亡／解放日期、返回日期、返回地點) 等。為整合這些記錄，該計畫建置猶太人大屠殺領域知識本體 (the Shoah domain ontology)，並採用 FOAF、Biographical Vocabulary、Dublin Core、Geonames 等語彙標準，以「人物」和「迫害事件」作為知識本體的主軸，再將人名、拘留地點等與其他外部資料庫如：DBpedia、VIAF、Geonames、dati.camera 等語意鏈結，串連人物及家庭關係、相關地點，呈現更清晰的猶太迫害史 (Open Memory Project, 2015)。

博物館社群的 LOD 計畫主要包含兩種類型：(一) 為了不同機構藏品之間資訊互通目的，例如：日本 LODAC Museum (Linked Open Data for Academia (LODAC) 計畫，以及美國的耶魯大學英國藝術中心 (Yale Center for British Art) 計畫。LODAC 計畫的主要目的是為

推動日本國內藝術領域的學術資料結構之創新，有鑑於日本國內的館藏機構皆已逐步數位化藏品，但缺乏一套知識本體架構，使之得以語意化，因此 LODAC 計畫援引國際通用之文化資產領域知識本體架構標準 CIDOC CRM (International Committee for Documentation, CIDOC; Conceptual Reference Model, CRM)，再將日本各公立藝文館藏機構（如：京都國立近代美術館、奈良國立博物館等）之藏品資料藉由此知識本體轉化成資源描述架構。因著此計畫，分散在日本國內各公立美術館與博物館的藏品與藝術家資料得以相互串連，同時促進在地藝術資源的交流 (Ontotext, n.d.)。Yale Center for British Art 計畫則是由大英博物館、耶魯英國藝術中心兩個文化機構共同合作，致力於藉由 LOD 傳播其藏品，但有鑑於這兩個機構的線上檢索系統的資料格式截然不同亦不容易整合，因此轉化成 LOD 的資料形式以滿足更多的使用情境，同時達到資料再利用，該計畫採用 CIDOC CRM 的知識本體架構，以滿足多種類的文化資產屬性，諸如手工藝品、博物館收藏、藝廊、考古學發現等，並可以讓資料的格式更一致，再利用性更高 (Ontotext AD, n.d.)。(二) 以博物館已建置的主題網站之資料為基礎的知識化工程，例如：英國 OpenART (Open Metadata for Art Research at the Tate) 計畫針對泰德美術館「The London Art World 1660-1735」網站既有檔案作為研究案例，進行資料結構的調整，達到資料語意化與開放鏈結資料之目標 (Allinson, 2012)。

以人時事地物為主軸之設計

LOD 計畫常藉由人、時、事、地、物等其中之一或組合為主軸進行設計，以突顯計畫的特色。人物方面，以上述提及的 Linked Jazz 及 Open Memory Project 等計畫為代表。時間方面，以上海圖書館 LOD 計畫所發展的中國歷史紀年表轉換為西元紀年表的知識本體為代表 (夏翠娟等人, 2012)。

事件方面，以 OpenART 計畫為代表，該計畫將研究範疇限縮在「藝術銷售」，並以「事件導向」為核心概念設計資料結構，同時在資料庫屬性設計之際力求呼應藝術研究需求。該計畫以每一筆藝術品的銷售作為「事件」指涉的對象，並透過特定「時間」(如：西元 1660-1775 年)與「地點」(如：倫敦)追溯此「事件」發生之歷程。每筆「銷售記錄 (sales)」由特定人物、地點以及資料來源 (包含報紙、財產清單、商品目錄等)記載的資訊所組成，並將之轉化為 RDF 格式 (Allinson, 2012)。

地點方面，以英國 PELAGIOS 鏈結古地理資料計畫為代表，其為多國學者合作計畫，藉由導入 LOD 技術，並透過數位文件、地圖資料的語意註解 (Semantic Annotation)，聚集不同用途的標籤，從中判讀資源之間的鏈結關係，使系統自動串連線上各種異質資料，諸如文本、博物館藏品、考古地點等與地理相關的線上資源，以幫助數位人文學者的資料能夠更易於被發現。該計畫在線上地名辭典 (Pleiades) 和數位歷史資源間建立大量連結，藉此方式存取資料，再由第三方進行分析或建置加值功能獲取資料。

物件方面，以 Graphity: The Danish Newspapers 丹麥新聞報典藏計畫為代表，該案例的資料來自丹麥國立大學圖書館，是一份歷史悠久且包括許多歷史資料的報紙，藉由將現存的三份報刊（Danish Newspapers 1634~1847、1848~1917、1918~1991）數位化之際，同時結構化及編輯現有資料與添加新資料，以讓使用者透過行動裝置的相容格式，簡單地發佈鏈結資料的網路應用程式，並具有多項功能，如互動式地圖等。該計畫採用 The Bibliographic Ontology、Time Ontology、The Places Ontology、Basic Geo、SKOS、Dublin Core Terms 等知識本體及 SIOC、FOAF 語彙集建置平台領域知識模型，再透過 LOD 技術將能整合各種單獨存在的資料成為統一的 RDF 格式，實現從印刷出版品到 RDF 資料的轉換，以及與檔案館之間的資料傳遞，並快速地提供各種來源的資料予搜尋引擎，進而優化搜尋結果。

研究設計與實施

研究對象

本研究以臺灣近代美術史的重要畫家陳澄波先生（1895~1947）的數位典藏資料為核心，包括畫作及檔案資料，諸如照片、書信、手稿、剪報、明信片等。此批資料經中央研究院臺灣史研究所檔案館進行數位典藏，依檔案的全宗理論及檔案描述編碼格式（Encoded Archival Description, EAD）標準予以組織編排，提供研究者瀏覽藏品資訊。爾後，中央研究院數位文化中心以該批檔案資料為基礎，擇取畫家不同時期的重要作品及相關文書檔案，與藝術史研究者合作，發展「從北緯 23.5°出發：陳澄波」主題網站（<http://chenchengpo.ascdc.sinica.edu.tw/>），以陳澄波先生的個人生平、藝術成就、環境變遷、社會關係等四大面向為主軸，以及十個主題串連他的人生故事，包含人生大事篇、畫作精選篇、帝展篇、家庭篇、日本篇、嘉義篇、上海篇、寫生旅行篇、畫友篇，以及畫壇組織篇等。該主題網站呈現與畫家相關的人、事、時、地、物之間的相關性，如配合事件或畫作，動態呈現多層次的時間軸，以及畫家群之間的社會網絡關係。此外，該網站也結合地理資訊系統（GIS），呈現百年來的歷史地圖。例如當在系統展示一幅畫作時，觀者同時可從地圖中知悉作品描繪的地點（如西湖），以及當時創作者的職涯經歷（如受聘於上海新華藝術大學教授美術）。或者，就畫作中的各種主題進一步探索相關所指，如吊橋、教堂塔樓、肖像畫等，再連結到外部系統「藝術與建築索引典」（AAT），以探索某個特定主題的範圍註及更多參考資訊，或者從更大群的主題標籤雲，點選任何一個使用者感興趣的關鍵詞。該主題網站以人工方式，經過文獻的爬梳與確認，將畫家之間的人際關係類型逐一標示出。本研究從上述資料取樣，包括相關作品、人物、組織及事件等類型的資料，進行鏈結資料轉化與資料補充建構。

研究設計

本文以 Tim Berners-Lee (2006) 提出的鏈結資料設計原則，以 RDF (World Wide Web Consortium, 2014a) 為理論基礎；再根據 W3C 組織提出的鏈結資料發佈之最佳實踐 (Best Practices for Publishing Linked Data) (World Wide Web Consortium, 2014b) 及網路發佈資料之最佳實踐 (Data on the Web Best Practices, DWBP) (World Wide Web Consortium, 2017)，以及史丹佛大學知識系統實驗室提出的知識本體發展方法 (Ontology development 101) (Noy & McGuinness, 2001) 等作為研究實作方法。以下就本研究的步驟與實施進行說明。

一、研擬使用案例

本研究以該主題網站為範疇，蒐集進階的藝術教育與初階的藝術研究過程中，使用者可能的提問，來源包括陳澄波相關研究文獻之內容分析，及訪問藝術學習者的問題清單，整理並歸納幾種常見問題類型，再請兩位藝術史研究者排序優先性而形成本研究的使用案例，主要包括：(1) 藝術家有哪些以某地點為場景的不同創作階段之作品及其相關檔案？其中入選某事件的是哪幾幅？(2) 藝術家在某組織可能的人際網絡關係？(3) 藝術家旅居過哪些地點？及在某地的居住時間？期間參與哪些活動或組織？藉由問題的查詢與提出，察覺並釐清人物、地點、事件、作品資訊之間的交互錯雜關係。

二、提取資料集

本文以「從北緯 23.5°出發 陳澄波」(以下簡稱：陳澄波主題網站) 的資料為核心，並取得其原始資料。限於時間與人力，本研究選擇以陳澄波本身的資訊為主，但為能完整回應前述的使用案例之提問，本研究經由五位臺灣近代美術史研究者的推薦，再經交叉篩選後，選出具代表性的研究資料作為延伸並補充相關的內容。綜言之，本研究的資料集內容主要取自：(A)「從北緯 23.5°出發 陳澄波」主題網站，並以(B)顏娟英，中央研究院歷史語言研究所「南國美術殿堂—臺灣美術展覽會(1927-1943)作品資料庫」網站，(C)顏娟英，《臺灣近代美術大事年表》(臺北市：雄獅圖書，1998)，(D)李淑珠，《表現出時代的「Something」：陳澄波繪畫考》，臺北市：典藏藝術家庭，2012 資料內容為補充(請詳表 1)。

表 1 本研究資料集類型與來源

資料集類型		資料內容之參考來源
人物		A, B, C
地點		A
組織		A, C
物件	繪畫	A, B, D
	照片	A, D
	明信片	A, D

(續表 1)

資料集類型		資料內容之參考來源
事件	展覽	A, B
	人物事件	A
	作品事件	A
分類與書目	參考書目	A
	展覽類型	B
	主題詞彙	A

上述涉及人、時、事、地、物不同方面的資料，經由挑選出最核心的事物特徵進行描述，在語意與機器計算下，透過最核心的屬性，可以延展出其他隱含的關係，進而建立更豐富的語意關係，實現智慧型的知識探索。

三、建立資料模型與知識本體

在鏈結資料的建構過程，知識本體讓我們得以清楚界定資料之間的脈絡關係，以及規範概念、屬性和實例的定義域（domain）與值域（range），是界定知識領域的基石，也是後續將數位資料發展為開放型 LOD 資料的基礎。RDF 是鏈結資料在建構知識本體最基本而關鍵的資料模型，其核心原理是基於圖資料模型，以及運用三元組（triples）的集合作為運算式的基本結構，包括主體／類別（subject）、謂詞／屬性（predicate）、客體／屬性值（object）來描述任何資源，成為意義完整而明確的描述（statement），讓電腦得以理解並處理描述的對象等訊息。在完成資料的三元組後，進入領域模型化的研究，經由界定重要概念，以及它們的屬性與關係，擷取出知識本體的概念結構，並關注於以形式化表達出概念的描述及其屬性。資料模型化的研究與分析，需能指出資料細節，描述物件中的關係。本研究的知識本體中之人物、事件、物件、地點、組織等主要概念模型之間的語意關係如圖 1 所示。其中，人物與物件之間藉由「創造者」屬性而建立語意關係，以便於使用者可以從「人物」的類別進入「物件」類別（詳圖 2），再藉由此類別的屬性探索更多語意關係的資源。

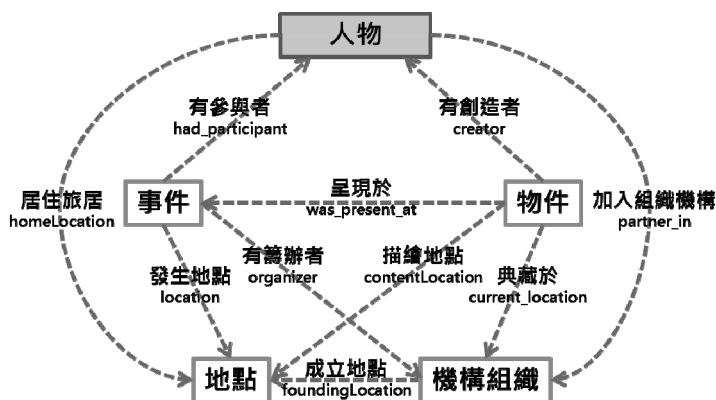


圖 1 本研究知識本體中人物、事件、物件、地點、組織等主要概念模型

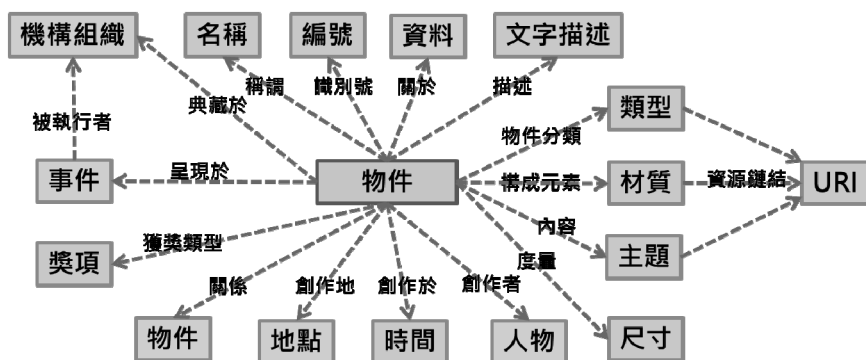


圖 2 本研究知識本體中物件為主之概念模型

圖 2 以「物件」類別為焦點的相關屬性及其屬性值，可能包含橫跨其它類別的屬性值（如地點、事物等），或者相同類別之間的屬性值（如材質、主題等）。再以「組織」類型為例，包括以陳澄波為中心的藝術家曾經參與的畫壇組織、就讀或任職學校與機構。組織機構是作家一生經歷中重要的組成部分，通過組織機構的鏈結資料，可以讓系統自動挖掘潛在的資源之間關係，例如：藉由人物就讀某特定學校事件之建立，帶出事件發生起始點，機器即可初步推論出臺灣前輩畫家間之特定校友關係（詳圖 3）。

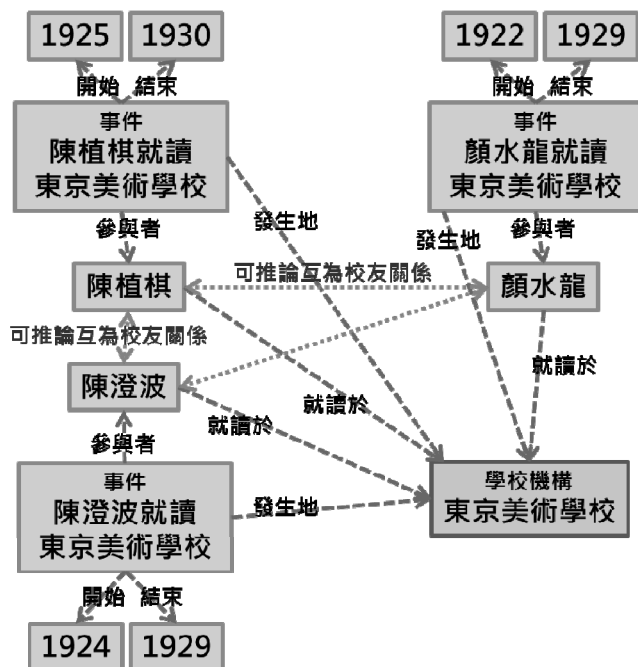


圖 3 本研究知識本體中人物間人際關係之推論示意：以陳澄波就讀東京美術學校時期之校友關係為例

四、採用可再利用的標準語彙及知識本體

為了將上述這些概念與關係，全面性地以標準詞彙進行描述，此階段的分析工作，旨在識別出適當的語彙集以作為可再利用的語意來源。本研究從具權威性及成熟的組織所發布之通用型知識本體開始調查與分析，為了識別出最合適的詞彙集來源，本計畫從開放知識基金會（Open Knowledge Foundation, OKF）維護的鏈結開放詞彙（Linked Open Vocabularies, LOV）提供之詞彙集清單進行評估與選擇。LOV 的主要目的是匯集詞彙的類別和屬性之定義，並以 RDF 或相關格式呈現（如 RDFS 或 OWL），來描述特定類型的事物。這些知識本體描述共通性的實體，諸如人、組織、事件或地理位置，經常作為概念模型設計的參考詞彙，藉由與其他來源詞彙的鏈結，增加本身知識本體的子類和子屬性，進而深化資料的脈絡。本研究採用已經釋放為 LOD 的控制詞彙（請詳表 2），包括：Schema.org vocabulary、CIDOC-CRM、Getty Vocabularies Program ontology、RDF Schema、Simple Knowledge Organization System (SKOS)、DCMI Metadata Terms、DBpedia 等共七套詞彙標準。

表 2 本研究建置知識本體所引用的標準語彙

前置詞 Prefix	命名空間 Namespace	語彙名稱 Vocabulary Title
schema.org	http://schema.org/	Schema.org vocabulary
crm	http://purl.org/NET/cidoc-crm/core#	CIDOC Conceptual Reference Model (CIDOC-CRM)
gvp	http://vocab.getty.edu/ontology#	Getty Vocabulary Program ontology
rdfs	http://www.w3.org/2000/01/rdf-schema#	RDF Schema
skos	http://www.w3.org/2004/02/skos/core#	Simple Knowledge Organization System (SKOS)
dct	http://purl.org/dc/elements/1.1/	DCMI Metadata Terms
dbpprop	http://dbpedia.org/property/	DBpedia

上述語彙標準主要是作為本文知識本體的各種屬性，選取自人物、事件、物件、地點、組織等主要類別的部分屬性作為例證，如表 3 所示。

表 3 本研究知識本體中各類別採用之屬性列舉

類別 Class	屬性 Property	注釋 Note
人物 Person	gvp:ulan1313_partner_in	描述人物參與的畫會、社團等組織
	schema.org:nationality	表示人物所屬的國籍
	skos:altLabel	表示人物的非偏好名
事件 Event	crm:P108_has_produced	描述創作事件所促成的實體物件
	dbpprop:years	描述展覽事件舉行之屆數
	schema.org:organizer	表示事件籌辦單位
地點 Place	rdfs:seeAlso	表示與地點相關之外部鏈結資料
	schema.org:hasMap	以地圖表示地點位置的 URL 連結
	schema.org:latitude	表示地點的緯度

(續表 3)

類別 Class	屬性 Property	注釋 Note
物件 Object	crm:P62 depicts	表示物件作品中描繪的事物
	dc:format	描述物件的尺寸
	schema.org:award	描述物件作品參加某展覽事件展出中獲得之獎項、殊榮
組織 Organization	crm:P1_is_identified_by	以編號表示本體中，組織機構實例之身分
	schema.org:foundingLocation	表示組織、機構成立地點
	schema.org:replacer	表示組織、機構延續、替代單位

藉由這些分析工作，可將標準詞彙進一步以三式的 RDF 資源描述框架，來描述每一小段資源。例如「陳澄波－就學於－東京美術學校」，其中，「就學於」，採用 Getty 人名索引典（ULAN）的標準詞彙，再如於〈清流〉畫作中所描繪的「雪」，採用 Getty AAT 索引典中的標準詞彙中「雪」（snow）的概念，再利用 CIDCO-CRM 的屬性詞彙，讓每一個語彙，都可以被清楚、正規化地表達後，方便機器的理解、計算與推演。

五、對應與轉換為鏈結資料

本研究分別從主題網站的資料集中抽取而成三元組敘述（statement），並依需求建立新資料以補充原資料集所不足的內容。在資料轉換的過程中，設計如下幾項重要研究步驟：

- （1）建立主體（subject）：識別出需要建立的主體，並進行概念層次的分類，以及為每個主體資源建立統一資源標誌符（URI）。
- （2）建立謂詞（predicate）：找出已建立的詮釋資料之資料集（如人物、物件等不同類型），建立屬性的對應規則，將原來的欄位從可再利用的語彙標準之中找出並對應到最接近語意距離的屬性。例如人物資料集的詮釋資料欄位「任職機構」，可以對應到 Schema.org 詞彙集的其中一個屬性「schema.org:worksFor」。
- （3）建立客體（object）：識別出需要建立的客體，判斷是「字串」或「事物」類型。若為前者，即可以直接採用原詮釋資料的資料值；若為後者，則需識別是否直接利用已有的語彙標準，以提供較客體資源更豐富的屬性。例如「畫作〈清流〉所描繪的內容所在地位於杭州」，此例之客體為「杭州」，本研究直接採用 Getty Vocabulary Program 知識本體的語彙，並對應至其 URI「<http://vocab.getty.edu/tgn/7001806>」，以提供使用者探索更多有關杭州的其他屬性。

研究結果與討論

本文探討鏈結資料在數位典藏資料之可能的效益，並以人物類型為焦點進行研究實作與討論。在整個研究實驗的過程中，總共發展出 30 個類別，及 57 個屬性以清楚指出資源之間

的語意關係、1,840 筆實例資料，產生總計 17,296 條三元組敘述。在研究實驗的整個步驟完成後，以 CC0 資料公開授權模式將這批資料開放，讓使用者及機器可處理的方式進行語意式查詢（SPARQL），以及資料下載（SPARQL endpoint）。回應本文的研究問題，包括鏈結資料如何強化數位典藏資源的探索性？鏈結資料應用在數位典藏資源，可能提供的脈絡化方式？鏈結資料如何自動推理產生具語意關係的資料？主要研究結果說明如下：

以鏈結資料強化數位典藏資源的探索性

從無法再延伸的字串（string）轉為可再延伸的事物（thing），是鏈結資料能夠強化傳統數位典藏系統的主要特色之一。傳統的資料庫，諸如大多數圖書館的書目資料庫或名稱權威檔，每筆資源記錄項下的屬性，包括作者，出版地等皆是視為字串（string）的概念。字串是由字元組成的序列，在系統中代表文字的資料類型，因此使用者無法再從文字類型的內容，進一步延伸探索該內容的更多特徵。本文藉由鏈結資料原理，除了將資料轉換為以 RDF 為基礎的模型外，並將構成最基本單元的每條資源之主體，謂詞及客體皆採用具有命名空間的 URIs，並讓主體與客體的資源屬性再以 RDF 模型表示，因而從（A）傳統上的字串內容，轉換為（B）可供延伸探索的事物（詳圖 4）。

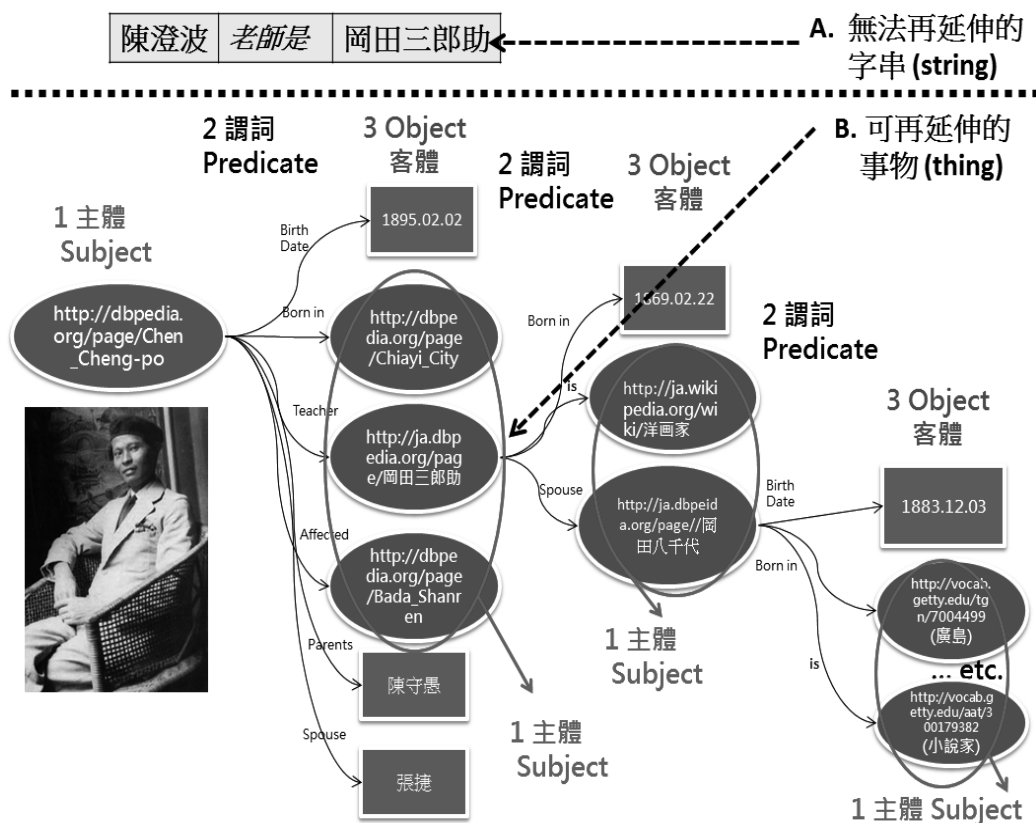


圖 4 無法再延伸的字串（string）與可再延伸的事物（thing）之示意圖

本研究結果可以提供使用者多層面 (faceted) 方式搜尋與瀏覽數位藏品, 分別包括人物、物件、組織機構、展覽、地點、時間等六個層面, 例如從「地點」層面, 可以繼續再由不同層面展現此資料集有關某特定地點的資源探索。以臺北為例, 根據本研究結果, 包括分別從「出生」、「逝世」、「成立組織」、「相關藏品」、「展覽」等層面將資料集內與地點「臺北」相關的資料展現搜尋結果 (詳圖 5)。

臺北

看看那些事發生在 臺北

出生
倪蔣懷 (1894-08-12)、立石達臣 (1905-03-11)

逝世
廖繼春 (1976-02-13)、李梅樹 (1983-02-06)、林玉山 (2004-08-20)、王白淵 (1965-10-03)、蕭添生 (1996-05-31)、陳進 (1998-03-27)

成立組織
中國文化大學、中國文化學院、國立歷史博物館美術審議、國立臺北教育大學、國立臺灣師範大學、太稻埕公學校、實踐大學、木蘭高等小學、橫山小學校、私立實踐家政專科學校、臺北公會堂、臺北州立臺北第三高等女學校、臺北市立中山女子高級中學、臺北市立福星國民小學、臺北市立美術館、臺

相關藏品
水源地附近 (陳澄波)、第一回臺灣省美展開幕部分審查委員合影、臺北東門 (陳澄波)、遠望公賣局 (陳澄波)、陳澄波與畫友合影於教育會館外

展覽
臺灣省全省美術展覽會_第3屆



圖 5 Linked Taiwan Artists : 臺北

資料來源：Linked Taiwan Artists 鏈結資料實驗系統 (<http://linkedart.ascdc.tw/lod/Place/PL0028>)

相較於目前藝術家相關數位典藏主題網站, 本結果提供以地點作為關注焦點, 聚集所有資料集內與特定地點相關的資料, 並有別於傳統全文式搜尋方法, 而是能依照資料的不同脈絡, 諸如: 聚集所有在臺北出生的藝術家 (如: 倪蔣懷)、在臺北舉辦的展覽會 (如: 臺灣省全省美術展覽會_第3屆), 或是以臺北為描繪場景的藝術作品或相關藏品 (如: 水源地附近、陳澄波與畫友合影於教育會館外) 等。提供使用者以某一個地點為核心的全貌觀 (holistic view), 再從任何感興趣的層面進一步探索, 例如: 進入「陳澄波與畫友合影於教育會館外」照片的資料, 可以繼續探索照片內的其中一位關係人「李梅樹」, 由此可以再以該藝術家為起始點, 分別探索其「相關藏品」、「參與展覽」、「參與組織」、「就讀學校」、「任職機構」、「老師」與「親屬」等。換言之, 本研究藉由鏈結資料方法, 將屬性資料值從無法再延伸的字串 (string) 轉為可再延伸的事物 (thing), 提供使用者可以不斷地從感興趣的「事物」以起始點再次探索相關資源。

所謂資源的再探索，在本文中包含不同的層次：(1) 相同資料集的不同屬性探索，例如：從一幅作品的「類型」或「材質」等不同屬性，探索資料集的相同材質之其他作品。此類探索功能可謂最常被應用的方式，最經典的案例包括：以鏈結資料方法為基礎的大英博物館線上藏品搜尋功能；(2) 從本身資料集連結至外部資料集，以近用更豐富的資料。例如從事件「芝加哥萬國博覽會」連結到 DBpedia，探索更多有關該展覽會的資訊、從一幅作品的主題「船」連結到 AAT；或者從一幅作品的人名與組織機構連結到 VIAF 等，請詳圖 6。

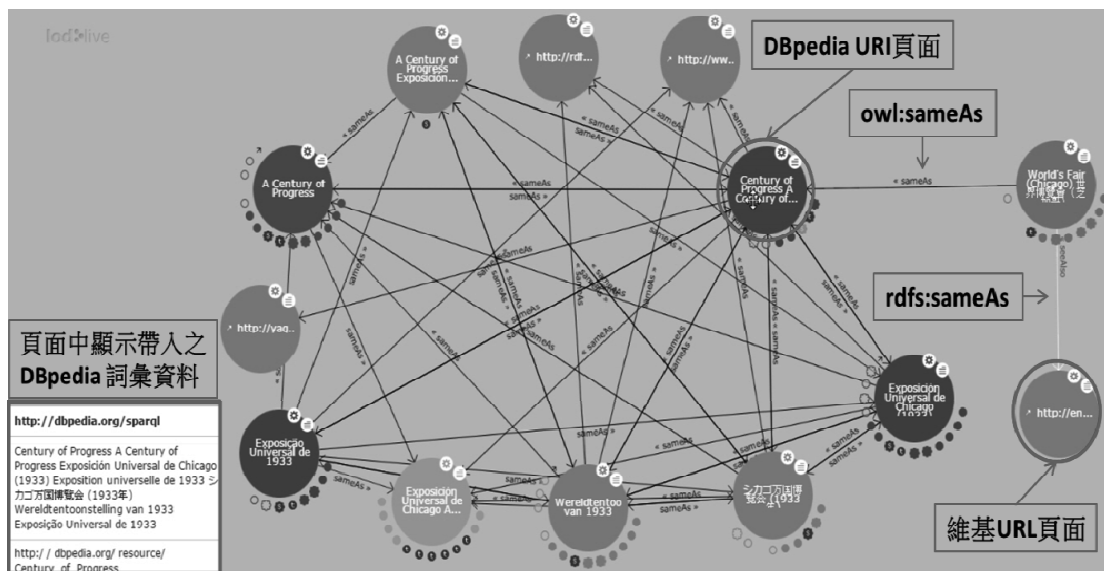


圖 6 鏈結外部 LOD 資料集

以鏈結資料提供數位典藏資源的脈絡化功能

脈絡 (Context) 有各種不同涵義，根據教育大辭書 (陳淑絹，2000)，其中一個意涵是認知心理學家在說明人的認知過程時的常用語。廣義言之，不僅包括人們在理解時，語言材料本身的上下文關係，也包括人們在認知客觀事物時，對客觀事物所具有的前後、上下、左右等各種背景關係。認知心理學強調脈絡對於人們的知覺、記憶、語言理解等活動有重要的影響，各種不同背景關係甚至可決定人們對於意義之間之理解。本研究的「脈絡化」則意指一個事物 (thing) 資源，以及其可再延伸的各種關係與事物或字串等資源。請詳下頁圖 7，是藉由 LOD 視覺化工具 LODLive，呈現經由用於 RDF 的查詢語言 SPARQL 之結果，提供瀏覽每項事物資源的脈絡關係。

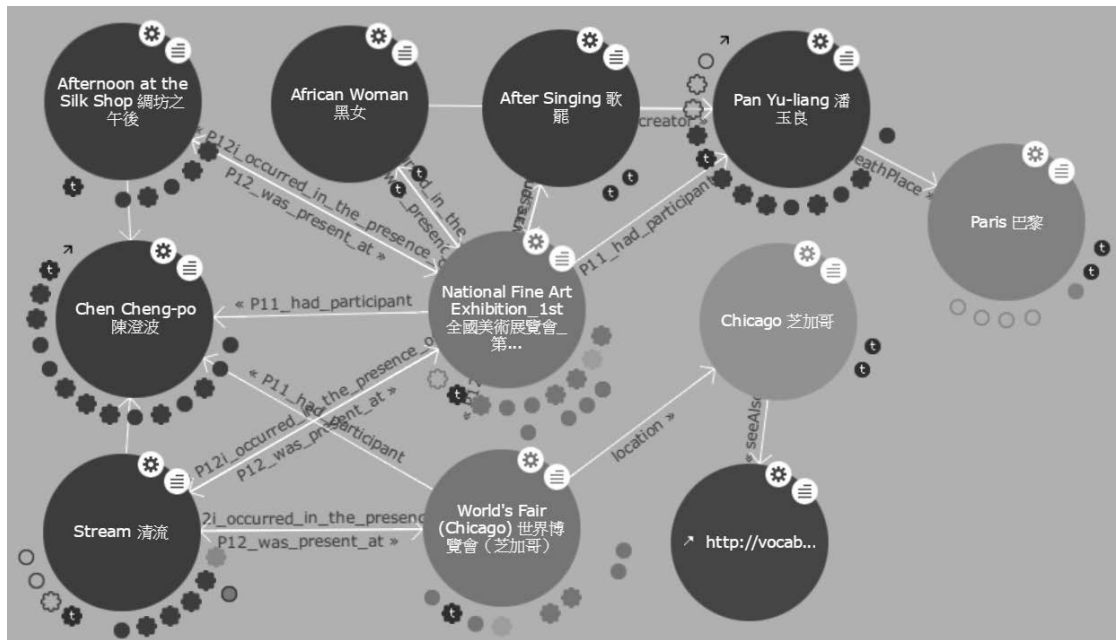


圖 7 可供機器處理的脈絡化研究材料

資料來源：Linked Taiwan Artists 鏈結資料實驗系統

(<http://linkedart.ascdc.tw/lod/Exhibition/EX0001-016#lodlive-section>)

圖 7 展示畫家陳澄波的作品〈清流〉與〈網坊之午後〉同時於第一回全國美術展覽會參展，而陳澄波在上海任教時期的同事潘玉良也有作品〈黑女〉與〈歌罷〉等入選於該回展覽會。此外〈清流〉同時入選於第一回全國美術展覽會及芝加哥世界博覽會。換言之，上述的脈絡資訊，可以提供使用者從一項本身感興趣的焦點，諸如「事件」再探索其相關的「物件」、「人物」及「地點」等關係。未來若將資料完整建立，則可以提供使用者探究更深入的提問，如「某個展覽會（如：臺灣美術展覽會），某國籍（如：台籍）畫家入選，有關某題材（如：自畫像）的作品有哪些？分別是哪些畫家？是屬於展覽會的哪個部門（如：西畫部）？這些畫家的專業訓練背景為何？」此類具深度探索性質的提問，在傳統的數位典藏系統，是相當不容易達到的功能。

以鏈結資料自動推理具語意關係的資料

本研究以 RDF 形式表達研究範圍的知識本體，藉由清楚界定的概念關係，以及推理的邏輯規則，進而讓系統有能力進行資料之間語意關係的自動推理。以人物為例，下頁圖 8 展示以自動推理為基礎的人際關係。

陳澄波

* 斜體字表示可能具有的社會關係。

其他名稱

Chen Cheng-po、Chen Cheng Po、Chen Chengbo、Chen Qinglan、Cheng-po chen、Tan Ting-pho、陳慶瀾

生卒時間

1895-02-02~1947-03-25

描述

1895年出生於嘉義市。1913年入臺灣總督府國語學校，師事石川欽一郎學習西洋水彩技法，1924年負笈日本，於東京美術學校（今東京藝術大學）圖畫師範科接受油畫家田邊至指導，又私

展開

相關作品

上海碼頭、不忍池畔(入選第22屆光風會展入選第3屆一九三〇協會展)、中國婦女裸體(入選)、二重橋、五里湖、人物(22)、人物(28)、人物速寫(172)-SB30(40.10-41.7)、人體(入選)

展開

參與展覽

一九三〇協會展_第3屆、上海藝苑繪畫研究所美術展覽會_第1屆、上海藝苑繪畫研究所美術展覽會_第2屆、世界博覽會(芝加哥)、中央美術展覽會_第8屆、光風會展_第15屆、光風會展_第

展開

就讀學校

嘉義公學校、本鄉繪畫研究所、東京美術學校、臺灣總督府國語學校

校友

伊原宇三郎、何德來、倪蔭懷、廖繼春、張秋海、李梅樹、李澤藩、林玉山、王白淵、田邊至、石川滋彥、范洪甲、藍綏鼎、陳植棋、鍾水龍、藍月桃甫、黃土水

老師

久米桂一郎、小林萬吾、岡田三郎助、淺井忠、田邊至、石川欽一郎、藤島武二、黑田清輝

學生

林玉山



圖 8 展示推測出的人際關係，不同類型（層面）資料之間的再延伸探索，提供不同脈絡

資料來源：Linked Taiwan Artists 鏈結資料實驗系統（<http://linkedart.ascdc.tw/lod/Person/PR0001>）

從圖 8 可以觀察到人物陳澄波的各種關係，其中在左下方，以斜體字標示的標題「校友」與「老師」與「會員」等是動態式資料，由系統自動推測出的人際關係。以校友關係為例，其方法並非直接建立人物之間的關係，而是藉由這些人物（如：陳澄波，林玉山）皆曾就讀於同一所學校（如：嘉義公校），進而推測他們可能為校友關係。不過，目前本研究僅採用最簡易的推論方式，尚未將時間因素納入推論，因此在此案例，由於有些人物（如：陳澄波）曾分別就讀與任教於同一所學校，但有些人物（如：林玉山）僅曾就讀同學校，因而會讓系統直接以「哪個人—就讀過—陳澄波—所就職過的組織」之概念作為查詢結果的轉換基礎，進而推論此二位人物具有校友與師生的關係。對於自動推論的資料結果，可能需要更小心檢視其在真實世界的正確性。以上述為例，本文經與陳澄波研究的學者專家確認，發現林玉山

與陳澄波確實是嘉義公校校友，但不是師生關係。因此，自動推論的結果如何呈現於系統介面，可適時提供提示性的語言，甚至容許使用者對於推論結果提供修訂建議，以精進資料的內容品質。

結論與建議

本文以人物及其相關數位典藏資源為對象，探索鏈結資料的理論與方法可以如何提升過去已建置的數位材料之主題網站。經由對於「從北緯 23.5° 出發：陳澄波」主題網站的材料進行語意化研究實作，結果顯示可以強化數位典藏資源的探索性、提供數位典藏資源的脈絡化功能，以及自動推理具語意關係的資料。此外，本文也梳理出可操作的研究實作方法，希望可助益於其他人物類型的數位典藏主題網站進行語意網典範轉移。基於此探索性的研究經驗與結果，建議未來可以更深入探討以下幾項議題：

強化資料之間語意關係的自動推論及結果的呈現方式

本文雖已展現資料間藉由自動推理而產生新的結果，但很顯然地由於本研究的實驗期間，在語意搜尋（Sparql query）過程中未考量到個人相關的起訖時間（如陳澄波在嘉義供學校的任教時間、林玉山在嘉義公學校就讀的時間）等屬性作為推論基礎，因此可能產生不正確的結果。建議，未來的語意搜尋除了需要納入更完整的屬性作為判斷關係的前提或描述限制之外，也需要建立推論結果的介面與提示，並讓使用者可以根據權威文獻或知識提出修正或補充的意見。

發展更完整的以人物為導向及其相關文物之資料模型

包括結合視覺化研究與工具，呈現藝術家不同層面、更細緻的人際網絡圈，如畫家陳澄波在日本時期與上海時期的朋友圈，以及畫作題材或風格是否與該時期的朋友圈之創作有關聯或影響。提供不同視野，讓使用者發現原先沒有期待發現的有意義之資料、事物或現象。因此，需要發展更為嚴謹的知識本體及推論規則。

發展資料清理與品質的方法論

由於數位典藏資源是過去十年來的成果，其基於當時的技術與標準，在語意化轉換時最大的挑戰之一便是資料品質與清理的問題（Zaveri et al., 2016）。

探索如何更全面地再利用已存在的控制詞彙作為豐富原有藏品的資訊

基於許多大型的文化資產組織，如美國國會圖書館或蓋堤研究中心皆已將過去數十年建置與維護的控制詞彙集以 LOD 發佈，此也意謂過去被視為高品質但卻也是高成本的資源集，藉由混搭與再利用方法，將有機會在語意網的環境發揮前所未有的效益。

探索與結合不同類型群眾外包（包含 Crowdsourcing 與 Nichesourcing 不同類型）的功能與機制

許多成功的研究案例，諸如 Linked Jazz、PELAGIOS 鏈結古地理資料計畫等均在研究過程中，嘗試以群眾外包模式讓使用者在某個階段加入合作，包括讓使用者補充更完整的資料，或者協助確認已建立的資料是否精確，藉由使用者參與而擴增知識本體（ontology）的設計，增強資料品質與完整度。由此，可以觀察出來這些「群眾」也可能來自專業或學術社群，因此建議可以視處理的資料與目標而採取不同的群眾外包類型。

致謝

本研究感謝科技部專題研究計畫（MOST 103-2420-H-001-009）補助，並感謝匿名審查委員的寶貴意見。另特別感謝中央研究院數位文化中心研究團隊的技術支援及協助資料整理。

附註

- [1]李澤藩素描創作手稿數位美術館計畫 http://203.74.112.6/asp/art_hall/index.htm
- [2]臺灣本土音樂家之影音典藏-李泰祥大師 <http://lth.e-lib.nctu.edu.tw/>
- [3]臺灣典範書家陳丁奇數位美術館 <http://museum02.digitalarchives.tw/teldap/2008/ChenDingChi/140.130.48.5/search/search.html>
- [4]打造福爾摩沙之聲：張連昌薩克斯風紀念館數位典藏計畫 <http://digicollection.nmns.edu.tw/saxophone/>
- [5]史惟亮教授音樂數位典藏計畫 <http://archive.music.ntnu.edu.tw/wlsh/>
- [6]臺灣民間藝術家數位藝術博物館之建置－以楊英風數位藝術博物館為例 <http://yuyuyang.e-lib.nctu.edu.tw/>
- [7]Online Coins of the Roman Empire, <http://numismatics.org/ocre/>
- [8]<http://summit2015.lodlam.net/2015/04/21/challenge-entry-graphity-the-danish-newspapers/>

參考文獻

- Abele, A., McCrae, J., Buiteiaar, P., Jentzsch, A., & Cyganiak, R. (2017). *Linking Open Data cloud diagram*. Retrieved from <http://lod-cloud.net/>.
- Allinson, J. (2012). Openart: Open metadata for art research at the tate. *Bulletin of the American Society for Information Science and Technology*, 38(3), 43-48.
- Berners-Lee, T. (2006). *Linked Data: design issues*. Retrieved from <https://www.w3.org/DesignIssues/LinkedData.html>
- Berners-Lee, T. (2009). *The next web*. Retrieved from https://www.ted.com/talks/tim_berners_lee_on_the_next_web

- Berners-Lee, T. (2010). *The year open data went worldwide*. Retrieved from https://www.ted.com/talks/tim_berners_lee_the_year_open_data_went_worldwide?language=en
- Bizer, C., Heath, T., Idehen, K. & Berners-Lee, T. (2008). *Linked data on the web (LDOW2008)*. In Proceedings of the 17th international conference on World Wide Web (pp. 1265-1266). ACM.
- Byrne, G., & Goddard, L. (2010). The strongest link: Libraries and linked data. *D-Lib magazine*, 16(11/12). Retrieved from <http://www.dlib.org/dlib/november10/byrne/11byrne.html>
- Durante, K., Weimer, K. H., & McGee, M. (2016, March). *Linked open data modeling for library cartographic resources*. In American Association of Geographers, 2016. Retrieved from https://scholarship.rice.edu/bitstream/handle/1911/88862/LD4P_AAG2016.pdf
- Illien, G. (2012, August). *Are you ready to dive in? A case for open data in national libraries*. In IFLA World Library and Information Congress, 78th IFLA General Conference and Assembly. Retrieved from <https://www.ifla.org/past-wlic/2012/181-illien-en.pdf>
- Jetzek, T., Avital, M., & Bjørn-Andersen, N. (2014). *Generating sustainable value from open data in a sharing society*. In International Working Conference on Transfer and Diffusion of IT (pp. 62-82). Springer, Berlin, Heidelberg.
- Kroeger, A. (2013). The road to BIBFRAME: The evolution of the idea of bibliographic transition into a post-MARC future. *Cataloging & Classification Quarterly*, 51(8), 873-890.
- Library of Congress. (2016). *Overview of the BIBFRAME 2.0 Model*. Retrieved from <https://www.loc.gov/bibframe/docs/bibframe2-model.html>
- Noy, N. F., & McGuinness, D. L. (2001). *Ontology development 101: A guide to creating your first ontology*. Retrieved from http://liris.cnrs.fr/alain.mille/enseignements/Ecole_Centrale/What%20is%20an%20ontology%20and%20why%20we%20need%20it.htm
- Ontotext. (n.d.). *LODAC museum: Providing Linked Open Data for academic resources*. Retrieved from <http://ontotext.com/knowledge-hub/case-studies/lodac-museum-linked-open-data-academia/>
- Open Memory Project. (2015). *Open Memory Project: Linked Data for the history of the Jews and the Shoah in Italy*. Retrieved from <http://summit2015.lodlam.net/2015/04/21/challenge-entry-open-memory-project/>
- Pattueli, M. C. (2012). Personal name vocabularies as Linked Open Data: A case study of Jazz artist names. *Journal of Information Science*, 38 (6), 558-565. doi: 10.1177/0165551512455989
- Ubaldi, B. (2013). *Open government data: Towards empirical analysis of open government data initiatives* (OECD Working Papers on Public Governance, No. 22). OECD Publishing.
- Voss, J. (2012). *Radically open cultural heritage data on the web*. Retrieved from http://www.museumsandtheweb.com/mw2012/papers/radically_open_cultural_heritage_data_on_the_w
- World Wide Web Consortium. (2014a). *Best practices for publishing linked data*. Retrieved from <https://www.w3.org/TR/ld-bp/>
- World Wide Web Consortium. (2014b). *RDF 1.1 concepts and abstract syntax*. Retrieved from <https://www.w3.org/TR/rdf11-concepts/>
- World Wide Web Consortium. (2017). *Data on the Web Best Practices*. Retrieved from <https://www.w3.org/TR/dwbp/>
- Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., & Auer, S. (2016). Quality assessment for linked data: A survey. *Semantic Web*, 7(1), 63-93.

Zuiderwijk, A., Janssen, M., van den Braak, S., & Charalabidis, Y. (2012). *Linking open data: challenges and solutions*. In Proceedings of the 13th Annual International Conference on Digital Government Research (pp. 304-305). ACM.

林信成 (2012)。資料模式 (data model)。圖書館學與資訊科學大辭典。檢自：<http://terms.naer.edu.tw/detail/1679034/>

【Lin, Sinn-Cheng (2012). Data model. *Encyclopedic Dictionary of Library and Information Science*. Retrieved from <http://terms.naer.edu.tw/detail/1679034/>】

夏翠娟、劉煒、趙亮、張春景、徐昊、朱雯晶 (2012)。關聯資料發布技術及其實現－以 Drupal 為例。中國圖書館學報，2012(1)，49-57。

【Xia Cui-Juan, Liu Wei, Zhao Liang, Zhang Chun-Jing, Xu Hao, & Zhu Wen-Jing (2012). The current technologies and tools for linked data: A case of Drupal. *Journal of Library Science in China*, 2012(1), 49-57.】

陳淑絹 (2000)。脈絡 (context)。教育大辭書。檢自：<http://terms.naer.edu.tw/detail/1308728/>

【Chen, Shu-Juan (2000). Context. *Encyclopedic Dictionary of Education*. Retrieved from <http://terms.naer.edu.tw/detail/1308728/>】

A Study of Linked Data for Digital Collections: A Case of the Painter Chen Cheng-Po

Shu-Jiun Chen

Assistant Research Fellow, Institute of History and Philology,
Academia Sinica, Taiwan (R.O.C.)
E-mail: sophy@sinica.edu.tw

Keywords: Chen Cheng-Po; Linked Data; Linked Open Data; Open Cultural Data; Semantic Digital Libraries;

【Abstract】

Linked Data enables digitized materials of cultural heritage institutions connect and integrate into Web of Data. The study focuses on people-based digitized materials. The datasets are selected from the “Starting out from 23.5°N: Chen Cheng-po” website as a case, to carry out Linked Open Data research and practices, demonstrating the context among people, organization, time, events, places, and artworks, and answering a complex questions based on semantic relations among resources. The results of the study enable machines to understand and process data, in order to facilitate access and re-use these research materials, so users can remix to create a new work, and thus enhance the use efficiency of digitized materials.

【Long Abstract】

Introduction

This study aims to explore how the thematic websites of digitized materials are converted into systems with semantic functions using Linked Data principles. Over the past 15 years (2002-2016), Taiwan e-Learning and Digital Archives Program (TELDAP), and the National Repository of Cultural Heritage (NRCH) have accumulated more than six million digitized materials from libraries, archives and museums (LAMs) in Taiwan. Due to the building time spans 15 years, the established digital resource systems remain web of documents paradigm which is based on hypertext navigation, instead of the examples of web of data. Therefore, it is difficult to explore information referenced and interconnected. The purpose of this study is to develop a Linked Open Data-based system to reflect the people and contextual

relationships among relevant people, time, thing, place, and object, answer complicated queries based on semantic relationships, provide visualization, and enable a large amount of data to be understood and processed by machines more conveniently to further facilitate the access and reuse of research data and provide users with efficacy of reassembling and further enhancing the digitalization of data. The research questions include:

1. How does Linked Data increase the discoverability of digitized materials?
2. Can Linked Data expose the context of digitized materials?
3. Can Linked Data perform automated reasoning to generate data with semantic relationships?

Research Design and Implementation

This study used the digitized materials of a renowned modern Taiwanese painter Chen Cheng-po (1895-1947), Chen's letters, diaries, postcards, manuscripts, photographs, and works of art. The materials were digitalized by the Academia Sinica. Moreover, the thematic bilingual website "Starting from 23.5° North Latitude: Chen Cheng-po" (<http://chenchengpo.ascdc.sinica.edu.tw/>) was developed. This study collected datasets from the mentioned above, including five types of datasets – relevant works, characters, organizations, and events, to perform Linked Data conversion and construction of data supplementation. The procedures and implementation of this study are explained as follows:

Exploring the cases to be used

During the data collection of advanced art education and elementary art research, for the possible questions to be raised by users, the inquiry and presentation of questions were used to perceive and clarify the complicated relationships among character, place, event, and work information.

Extracting the dataset

This study obtained the raw data, and selected the most central object characteristics for description. With the algorithms of semantics and machine, other hidden relationships were extended from the most central properties to further develop more abundant semantic relationships.

Developing the data model and domain ontology

The study of data modeling has to be able to point out the relationship between data details and described object. The domain ontology developed in this study included the semantic relationships of main conceptual models, such as character, event, object, place, and organization.

Adopting reusable vocabularies and ontologies

This study adopted a total of seven sets of vocabularies that were released as the controlled

vocabularies of the LOD, including: Schema.org vocabulary, CIDOC-CRM, Getty Vocabularies Program ontology, RDF Schema, Simple Knowledge Organization System (SKOS), DCMI Metadata Terms, and DBpedia.

Mapping to and converting to Linked Data

This study built triple statements from the datasets of thematic websites, respectively, including the analysis of subject, predicate, and object.

Research Results and Discussion

This study developed a total of 30 categories and 57 properties to clearly indicate the semantic relationships among resources. A total of 17,296 triples were generated from 1,840 metadata records. Upon completion of entire procedures of the research experiment, all these metadata are under the terms of the Creative Commons CC0 Universal Public Domain Dedication, and allow query Linked Data with SPARQL. The main results are as follows:

Using Linked Data to increase the discoverability of digitized materials

Converting a string that can no longer be extended into a thing that can be extended is one of the main characteristics of Linked Data to strengthen traditional digital archives system. The results can provide users with multifaceted methods to search and browse digital collections, including 6 dimensions – people, object, organization/institution, exhibition event, place, and time.

Using Linked Data to provide contextualization function of digital archive resources

The “contextualization” in this study refers to the resource of a thing, as well as resources, such as various relationships and things or strings that can be further extended. This study used LOD visualization tool, LODLive, to reflect the results of query language SPARQL used in RDF and provide the browsing of contextual relationships of each thing/resource.

Using Linked Data to perform automated reasoning of data with semantic relationships

This study used clearly defined conceptual relationship and logical rules of reasoning to further enable the system to perform the automated reasoning of semantic relationships among data. Taking people for example, Figure 8 shows the automated reasoning-based interpersonal relationship.

Chen Cheng-po

* *Italics* represent possible personal relationships of each related artist.

Also known as

陳澄波、Chen Cheng Po、Chen Chengbo、Chen Qinglan、Cheng-po chen、Tan Ting-pho、陳慶瀾

Date

1895-02-02~1947-03-25

Description

Chen Cheng-po was born in Chiayi City. He enrolled in the Taiwan Governor-General's National Language School in 1913, where he studied Western-style watercolor painting under Kinichiro Ishikawa. In 1924, he went to Japan to receive formal academic training in art under Japanese oil

►more

Work

[A Distant View of Guanvin Mountain](#) (selected for exhibition)、[A Distant View of the Monopoly Bureau](#)、[A Postcard from Chen Cheng-po to Chen Bi-nu_2](#)、[Afternoon at the Silk Shop](#) (selected for exhibition)、[Air Raid Drills](#) (selected for exhibition)、[Ali Mountain](#) (selected for exhibition)、

►more

Exhibition

[1930 Society Exhibition_3rd](#)、[Central Art Exhibition_8th](#)、[Fujian Provincial Art Exhibition](#)、[Hakujitsukai Exhibition_2nd](#)、[Hakujitsukai Exhibition_3rd](#)、[Hakujitsukai Exhibition_4th](#)、[Hakujitsukai Exhibition_5th](#)、[Hongo Art Exhibition_4th](#)、[Imperial Art Exhibition_10th](#)、[Imperial Art](#)

►more

Education

[Chiayi Public School](#)、[Hongo Painting Institute](#)、[Taiwan Governor-Generals National Language School](#)、[Tokyo School of Fine Arts](#)

Alumnus

[Chen Zhi-qi](#)、[Fan Hong-jia](#)、[He De-lai](#)、[Huang Tu-shui](#)、[Itaru Tanabe](#)、[Li Mei-shu](#)、[Li Ze-fan](#)、[Liao Chi-chun](#)、[Lin Yu-shan](#)、[Ni Chiang-huai](#)、[Ran In-ting](#)、[Shigehiko Ishikawa](#)、[Toho Shiotsuki](#)、[Usaburo Ihara](#)、[Wang Bai-yuan](#)、[Yan Shui-long](#)、[Zhang Qiu-hai](#)

Teacher

[Chū Asai](#)、[Itaru Tanabe](#)、[Keiichi Kume](#)、[Kinichiro Ishikawa](#)、[Mango Kobayashi](#)、[Saburosukey Okada](#)、[Seiki Kuroda](#)、[Takeji Fujishima](#)

Student

[Lin Yu-shan](#)

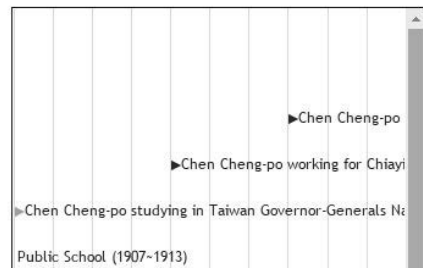
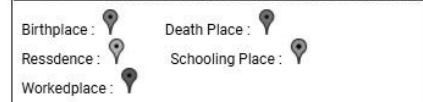


Figure 8 Display of Estimated Interpersonal Relationship; the further extended exploration among different types (dimensions) of data provided different contexts

Source: Linked Taiwan Artists Linked Data Experimental System (<http://linkedart.ascdc.tw/lod/Person/PR0001>)

Various relationships of the painter Chen Cheng-po can be shown in Figure 8. The italic titles in the lower left corner, “*Alumni*,” “*Teacher*,” and “*Member*,” were dynamic data, which were the interpersonal relationships automatically estimated by the system.

Conclusion and Suggestions

This study selected a painter and his relevant digital archive resources as the objects to explore how Linked Data principles can improve the thematic websites of digitized materials built in the past. The study performed a semantic investigation on materials of thematic website “Staring from 23.5° North Latitude: Chen Cheng-po.” The results showed that the discoverability and contextualization of digital archive resources could be strengthened, and the data with semantic relationships could be automatically reasoned. Moreover, the study has demonstrated workable research methods. Hopefully, these methods can be beneficial to the paradigm shift of Semantic Webs of digital archive thematic websites of other types of characters. Based on the experiences and results of this exploratory study, future studies are suggested to further probe into the following issues: (1) to strengthen the automatic reasoning of semantic relationships among data and presentation of results; (2) to develop more complete character-oriented data model and relevant cultural artifacts; (3) to develop the methodologies of data cleaning and quality; (4) to explore how to more comprehensively reuse existing controlled vocabularies as abundant information of original collections; (5) to explore and combine the functions and mechanisms of different types of crowdsourcing (including different types of crowdsourcing and nichesourcing).

【Romanization of Chinese references is offered in the paper.】