

# 探索數位典藏的詮釋資料與索引典之多語化(註 1)

## An Exploration of Developing Multilingual Metadata and Thesauri for Digital Libraries

陳 淑 君

Shu-Jiun Chen

中央研究院資訊科技創新研究中心研究助技師

Assistant Research Engineer, Research Center for Information Technology Innovation,  
Academia Sinica

E-mail: sophy@sinica.edu.tw

城 菁 汝

Ching-Ju Cheng

國立臺灣藝術大學藝術管理與文化政策研究所博士生

Ph.D. Students, Graduate School of Art Management & Culture Policy,  
National Taiwan University of Arts

E-mail: elizza29@gmail.com

陳 雪 華

Hsueh-Hua Chen

國立臺灣大學圖書資訊學系暨研究所教授

Professor, Department and Graduate Institute of Library and Information Science,  
National Taiwan University

E-mail: sherry@ntu.edu.tw

### 【摘要 Abstract】

本文以兩個數位典藏多語化國際合作計畫為研究案例，進行詮釋資料與索引典多語化方法架構之討論。首先，藉由索引典探討中文與英文控制詞彙語義互通的架構，提出翻譯、對應、在地化及創造等四個模組，其次，探討以中文詮釋資料為基礎，英文化的四項步驟，包括比對、揀選、翻譯及控制詞彙。最後，本文討論兩個案例藝術與建築索引典與 ARTstor 合作計畫所面臨的關鍵議題，包含中英文詞彙

對應與翻譯時無法互為完全等同關係的問題，及詮釋資料翻譯與詮釋資料品質問題，並提出解決方法與建議。

In this paper, we aim to explore methodologies of developing multilingual metadata and thesauri for digital archives through case studies of two research projects. We first analyze the Chinese thesaurus and the English thesaurus to construct a methodological framework for Chinese-English lexical and semantic interoperability, proposing four main modules, including translation, mapping, localization, and creation. The next step is to examine the four steps involved in transcribing the Chinese metadata into the English form, including mapping, selection, translation, and vocabulary control. This paper further looks into several key issues faced by the two research projects, such as the varying degrees of semantic equivalence between Chinese and English vocabulary, metadata translation, and the quality of metadata.

**關鍵詞** Keyword

多語化 詮釋資料 後設資料 索引典 數位典藏 數位圖書館

Multilinguality ; Metadata ; Thesauri ; Digital archives ; Digital libraries

## 壹、前言

自從美國國家科學基金會於 1994 年開始數位圖書館先導計畫以降，多語化成為全球各國數位圖書館研究領域裡最重要的議題之一(Borgman, 1997; Fox & Marchionini, 1998; Oard & Diekema, 1998; Zeng & Chan, 2004; Clough & Eleta, 2010)。台灣從 2002 年開始啟動大規模的數位典藏國家型計畫，在累積兩百萬件數位藏品之際(計畫辦公室，2007)，於 2007 年開始成立國際合作分項計畫，其中一項重要的任務是建立數位成果網站的多語化內容。數位典藏與數位學習國家型科技計畫的目標是永續經營台灣的數位文化資產，並且將計畫的成果推廣至產業界、教育界與學術界。此國家型計畫於 2012 年底已正式結束，產生超過六百萬件數位化的藏品，目前已整合於數位典藏聯合目錄，除了數位圖檔並建立以中文為主的詮釋資料(註 2)，公開於網際網路供全球檢索與利用(計畫辦公室，2013)。然而要將這些豐富而多樣的數位藏品進一步地與世界分享，極需克服語言隔閡的挑戰。本文係根據作者參與中央研究院及數位典藏與數位學習國家型科技計畫的多語化研究計畫之建構經驗，在數位圖書館的脈絡下提出索引典與詮釋資料多語化的研究框架，旨在探索並發展中文資訊組織的多語化議題與方法，終極目標是為讓使用者能以多種語言檢索數位化的中文數位典藏資源。首先，本文以數位典藏與數位學習國家型科技計畫與美國蓋提研究中心的國際合作計畫－中文版藝術與建築索引典(AAT Taiwan)為例，說明對應與翻譯三萬五千筆英文控制詞彙成為中文詞彙的方法與步驟。其次，本文呈現數位典藏與數位學習國家型科技計畫與美國美崙基金會(The Andrew W. Mellon Foundation)的 ARTstor 數位圖書館之合作計畫，分析如何為中華文化概念為基礎的數位藏品建置英文詮釋資料。最後，本文提出資訊組織多語化的議

題討論及結論。

## 貳、數位典藏多語化的研究方法與框架

數位典藏與數位學習國家型科技計畫(為行文方便，以下簡稱 TELDAP)自 2007 年開始，分別以三個途徑為數位化藏品提供多語化機制。第一個途徑，是成立 AAT Taiwan 研究團隊(<http://aat.teldap.tw/>)，並與美國蓋提研究中心合作，進行藝術與建築索引典(Art & Architecture Thesaurus, AAT)英文術語的中文文化研究。同時將 AAT-Taiwan 與數位典藏聯合目錄的百萬件數位藏品詮釋資料整合，主要目的是為了提供全球的使用者以在地語言檢索與瀏覽台灣的數位典藏資源。第二個途徑，是藉由與美國 ARTstor 的先導國際合作計畫，挑選出五百筆 TELDAP 數位化的藏品，並進行藏品詮釋資料的英文化研究。第三個途徑，是建立 TELDAP 成果的多語化網站(<http://culture.teldap.tw/culture/>)，粹取中文版成果入口網(<http://digitalarchives.tw/>)的重要典藏資源進行英文，日文與西班牙文翻譯(Chen & Chen, 2008)。限於篇幅與時間，本文將僅詳細說明上述的第一與第二種途徑之多語化研究方法與框架。

### 一、藝術與建築索引典中文文化研究 (AAT Taiwan)

美國蓋提研究中心從 1970 年代便開始發展藝術與建築索引典(AAT)，目前已建立三萬五千多筆概念詞彙及十三萬一千條多語及同義詞(Harpring, 2009)。該索引典是一部符合國際標準組織(ISO)規範，以階層式架構呈現的資料庫，其層級架構包括等同關係、階層關係及聯想關係。詞彙的範疇囊括藝術、建築、裝置、物質文化與質材的概念。自 2008 年開始，TELDAP 與美國蓋提研究中心共同合作研究開發中文版的藝術與建築索引典(AAT Taiwan)。

雖然國際標準組織(ISO) Guidelines for the establishment and development of multilingual thesauri 等標準提供諸多設計發展多語索引典的指引(ISO, 1985, 2011a, 2011b)，相關的研究文獻也提供許多的發現與議題(Hudon, 1997; Zeng & Chan, 2004; Landry, 2004; Park, 2007)。但是基於語言的獨特性，尤其是英文與中文是兩個距離相當遙遠的語系及文化脈絡，大規模而有系統地從英文索引典轉化為中文，尚只有很少的研究文獻(Liang & Sini, 2006)，甚至以藝術為主的中英雙語索引典研究與發展更是付之闕如。因此在最初規劃藝術與建築索引典中文版(以下簡稱 AAT Taiwan)時，研究團隊面臨諸多方法論與操作面的議題與抉擇，包括應該如何發展一套方法論？該使用什麼方式來翻譯？是否僅翻譯偏好詞彙？不同語言之間的轉換是翻譯還是對應(mapping)？兩者之間該如何區別及落實為可操作的工作任務？應該如何組成一個多語化團隊？完整的團隊成員需要由那些資格與背景組成？翻譯過程中，中英詞彙對應的原則是什麼？如

果有一對多的情況，該如何選擇描述語(或偏好詞彙)？若是遇到文化空缺的詞彙，該如何進行對譯？

為讓此項國際合作計畫更有效落實，TELDAP 與美國蓋提研究中心於 2008 年到 2013 年之間，舉辦了六場國際術語學工作小組(International Terminology Working Group, ITWG)會議，與同在執行 AAT 多語化的成員國家之研究機構與學者，包括美國(英文)、台灣(中文)、荷蘭(荷蘭文)、智利(西班牙文)、德國(德文)、瑞士、巴西及英國等進行跨國多邊研究討論會，共同探討術語多語化的理論與實踐，控制詞彙如何改善數位圖書館的搜尋與瀏覽功能等多項議題。歷年的相關探索題目，諸如多語化對應關係面臨的議題與挑戰：西班牙文版藝術與建築索引典翻譯計畫的經驗與中英兩種語言間等同關係對應的方法與問題等。藝術與建築索引典中文版(AAT Taiwan)所使用的研究方法與框架如圖 1，共涵蓋翻譯、對應、在地化和創造四個模組，逐一探討如下。

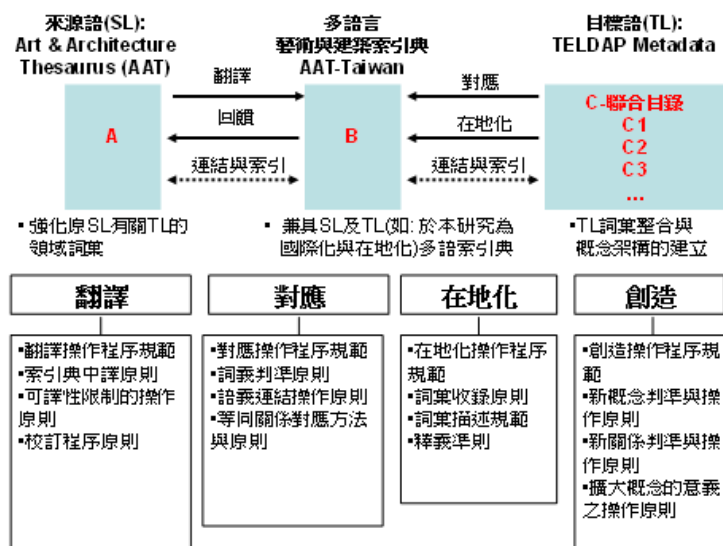


圖 1 多語化索引典研究方法與框架

### (一)翻譯模組

所謂翻譯在此意指將英文版的 AAT 翻譯成中文版，其中包括詞彙及範圍註兩項中文翻譯。例如一個有關藝術風格的 AAT 詞彙 Rococo，其範圍註為「Refers to that period primarily of decorative art that emerged in France ca. 1700 at the court of Louis XV, and dominated Europe until it was superseded by the Classical Revival in the late 18th century. The style is characterized by opulence, asymmetry, grace, gaiety, and a light palette of colors, in contrast to the heavier forms and darker colors of the Baroque.」，中譯後詞彙為洛可可，中譯的範圍註是「指約西元 1700 年，法王路易十五時期的裝飾藝術，後風行整個歐洲，直到十八世紀後期被新古典主義所取代。其特徵為富麗奢華、不對稱、優美雅致、歡樂與色彩明亮，與巴洛克厚重的形式和暗色調迥異。」從管理角度檢視，初期 AAT Taiwan 研究團隊在翻譯模組方面最重視三項任務，分別是徵才、教育訓練以及品質管理。在徵才方面，翻譯成員的背景盡量要求外語、翻譯及藝術等專業多元性。以 2010 年為例，十餘位翻譯成員當中 46% 主修翻譯，31% 主修外國語言與文學，15% 則有藝術與建築背景；校訂成員中 43% 主修藝術與建築，29% 主修外國語言與文學，15% 主修翻譯。教育訓練方面，需要教導譯者與校訂者遵循一致性規則及步驟來翻譯，同時研究團隊發展一系列翻譯操作程序規範、索引典中譯原則、可譯性限制的操作原則、校訂程序原則等，讓整個翻譯流程能進行的更流暢順利。例如：提供譯者及校訂者一份經研究分析後的固定中英詞彙翻譯表，指引譯者與校訂者在翻譯先導詞或一般性詞彙時的標準譯法。譯回內容，則需設計方法與步驟可以檢驗翻譯與校訂過後的範圍註，並且確認範圍註的精確度和流暢

度。在實際操作面，最常遇見的是曲解或是無法完整呈現原句意涵的情況，以及該詞彙在東西方概念下無法完全精準對應或是有對應錯誤的情形。此外，研究團隊也採用線上論壇模式回應譯者的問題，此舉可免除管理團隊重複回答相同問題的情況，亦可讓譯者互相學習並提升翻譯效率。

### (二)對應模組

對應(mapping)在此意指英文詞彙與中文詞彙之間的語義對譯，其可能包含兩種模式，第一種模式是將來源詞彙(如：在本研究是 AAT 的英文詞彙)對譯到目標詞彙(如：在本研究是概念相同的中文詞彙)，在此研究計畫的每條 AAT 英文詞彙皆需有至少三個具權威性來源的資料作為翻譯佐證，方被認為正式的中譯名稱，而非由譯者任意自行翻譯的詞彙。這些具權威性的翻譯來源，也許是權威機構使用的中文控制詞彙，或是由專家學者認可的參考資料中找出的中文詞彙。如果是常用字彙，則多以 TELDAP 聯合目錄作為語料庫基礎，尋找最多使用的中譯名稱。在執行對應模組時，翻譯團隊以具權威性的資料來辨別中英詞彙對應關係，並以五種對應關係來表示原英文詞彙與中譯詞彙之間的關係，分別是精準等同、不精準等同、部分等同、一對多等同以及不等同。

第二種模式，是反之將目標詞彙(如：在本研究是國立故宮博物院的中文詞彙)，對譯到來源詞彙(如：在本研究是 AAT 的英文詞彙)。根據 Chen & Chen (2012)以國立故宮博物院器物類的中文控制詞彙，作為中英詞彙概念對應分析的基礎，並以上述五種對應關係作為對應分析。整個完整中英詞彙對應流程包括如下重要步驟：(1)為中文原詞彙蒐集並辨識具權威性的英文詞彙(詳表 1)。(2)選定最合適的中文詞彙成為此概念在索引典中文版的偏好詞彙，其他詞彙為非偏好詞彙，但仍可被使用者檢索(詳表 2)。(3)選擇上述五種關係之一來代表目標

詞彙與來源詞彙之間的等同關係類型，以作為未來 基礎(詳表 3)。  
 檢索系統在處理概念檢索時的精確與模糊之設計

表 1  
 來源詞彙與目標詞彙語義對應查證之分析示例

目標詞彙 (取自故宮)	已存在的翻譯	參考資料	目標詞彙的參考來源
孔雀藍	Peacock blue	藝術名詞與技法辭典(2002)中文 版本	數位典藏子計畫國立故宮博物院 詮釋資料規格書第 1.2 版本， 第二十頁

表 2  
 目標詞彙對應至來源詞彙及相應結構之分析示例

目標詞彙 (取自故宮)	在目標詞彙的 概念脈絡	對應於的來源詞彙(AAT)及其概念脈絡、意義
繡花	繡花(〈技法〉)	Term: embroidering Hierarchy: Processes and Techniques Facet: Activities Facet Note: embroidering(needleworking(process) , <needleworking and needleworking techniques> , ... Processes and Techniques)

表 3  
 目標詞彙與來源詞彙的對應關係類型之分析示例

對應關係類型	在本研究的意涵	範 例	
		目標詞彙	來源詞彙
完全等同	目標詞彙(取自故宮)的中 文與來源詞彙(AAT)的英 文代表的概念完全相同	仰韶文化 (〈考古學文化〉)	Yangshao(<Chinese Neolithic periods> , <Chinese prehistoric periods> , ... Styles and Periods) ID: 300173481

(三)在地化模組

在地化模組，可以區分為兩種互為雙向流程。  
 第一種是從原為英文的來源詞彙，進行中文化而成

為目標詞彙的在地化過程。中文化的過程並非只是  
 單純地進行翻譯，由於涉及兩個不同文化背景及對  
 一個概念的指涉範疇之狹廣與深淺，在來源語與目  
 標語的轉譯之間，可能需要經過在地化的處理，以

讓一個概念在兩個語言的連結之際，獲得最佳化的語義互通性。此外，來源語在轉譯為目標語的過程，可能出現多種譯法，或是來源語較目標語具有更模糊的概念。例如：「ceramicists」(ID: 300025388)在英文意指「從事陶藝創作、陶瓷品製作或製陶相關技術研發者」。在轉譯為中文時，不同文獻及權威工具書可能將此詞彙對應到「陶匠」、「陶瓷藝術家」、「陶藝家」、「陶瓷工」及「製陶業者」等不同的中文詞彙。基於索引典的檢索功能，需要蒐集並聚合這些相對於來源語的目標語之所有同義詞或類同義詞。對於多語索引典而言，這是一種在地化的過程，因為中文化的詞彙已盡可能納入各種相對於來源語的在地詞彙。

第二種是將原為中文的目標詞彙，進行英文化而成為來源詞彙的一部分。此情況經常出現在：(1)當中文詞彙無法找到概念完全相等的英文詞彙時，在此研究計畫會建議修改 AAT 裡最近似的英文詞彙範圍註內容，希望能達到中英文概念的範疇完全等同的狀態，(2)或是需修訂 AAT 的層級架構，讓新加入的中文概念詞彙能夠融入整體架構裡。

#### (四)創造模組

在多語化的研究發展過程，若是以一個已存在的來源語索引典(如：AAT)為基礎，藉由翻譯或對應等方法建立目標語索引典(如：AAT-Taiwan)，會產生某些概念詞彙是部分等同或完全對應不到的

問題，其中若是主要形成的原因是文化或與語言差異性，則會需要利用創造模組，對來源語所缺乏的詞彙或結構進行從新的建立與發展，以讓來源詞彙與目標詞彙達到最佳化的多語索引典。例如，以西方藝術為基礎發展而成的 AAT，可能無法全面性包含其他文化中的重要語彙諸如中國藝術領域，因此可能造成目標詞彙在經過轉譯整個來源詞彙之後，依然不是最完整的詞彙集，甚至概念結構也會有所缺漏。以 AAT-Taiwan 的研究發展為例，可以建立三個解決的操作方式包括：(1)找出來源詞彙(如：AAT)沒有，但是對目標詞彙(如：中國藝術領域)具重要性的詞彙，(2)將這些中文詞彙進行英文轉譯，並且嘗試辨識出能夠讓這些詞彙整合在 AAT 適當的概念結構之中，(3)最後將這些 AAT-Taiwan 所新創造的詞彙貢獻於 AAT 中。以「中國書體」為例，「隸書」、「楷書」、「篆書」等詞彙皆為中國書畫領域的重要書體形式，然而在 2010 年的 AAT 版本並未收錄這整組詞彙，甚至概念結構也完全缺乏。經過研究分析，AAT-Taiwan 團隊與書法研究的專家學者共同發展一組中國書體的控制詞彙及語義結構，包括甲骨文、金文、篆書、隸書、楷書、行書、草書等不同概念詞彙，並借用 AAT 中〈阿拉伯書體〉層級向下囊括了十二種阿拉伯書體形式的原則與作法，增設〈中國書體〉層級到“書體(書寫)－〈依形式區分之書體〉”之下，同時成為「阿拉伯書體」同位階的架構，請參考下圖(詳圖 2)

## Example : Chinese Scripts

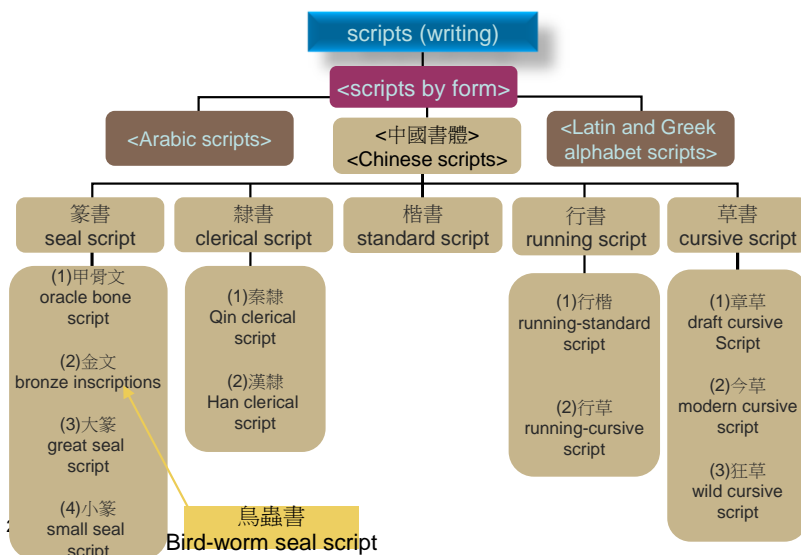


圖 2 創造新詞彙

## 二、數位典藏詮釋資料英文化研究 (TACP)

對於一個擁有百萬筆中文詮釋資料的數位典藏資源，如何發展英文化詮釋資料，以提供世界各地更多使用者近用數位資源是數位典藏推廣研究的一項挑戰。就策略方法，可以發展合集層級 (collection level) 或館藏層級 (item level) 的英文化詮釋資料。前者以一個資源合集為單位建立一筆詮釋資料，例如：故宮博物院書畫典藏資源為一個合集；後者則是以每項單件館藏為單位並逐一建立詮釋資料，例如：為故宮博物院書畫合集的其中一幅作品「宋郭熙早春圖」建立的詮釋資料，是屬於館藏層級。本文是以精選館藏層級的藏品之中文詮釋資料，與美國 ARTstor 合作進行英文化先導研究，主要目的是想探索來自不同典藏計畫，各有不同特色的中文詮釋資料，在與國際為平台的數位圖書館

整合時，轉譯為英文詮釋資料的方法與步驟，及其成本效益。

ARTstor 是由美國美崙基金會 (The Andrew W. Mellon Foundation) 成立的非營利數位圖像圖書館，收錄超過一百五十萬筆藝術、建築、人文和科學的圖像 (註 3)。圖像 (image) 資料主要是來自世界著名博物館、圖書館、大學、研究單位、相片檔案館、攝影師、學者、及藝術家的收藏。ARTstor 最大特色是提供高解析度數位影像，藉此幫助學者、教育工作者、和學生利用圖像資料進行研究、教學和學習。TELDAP 與 ARTstor 合作計畫 (TELDAP & ARTstor Collaboration Project，簡稱為 TACP) 起源於 2009 年 12 月，ARTstor 希望能收錄臺灣 TELDAP 豐富的數位圖像資料，尋求與 TELDAP 建立國際合作關係。TELDAP 評估此國際合作計畫能協助藏品詮釋資料英文化、提升臺灣數位圖像於國際平台的能見度、以及作為數位典藏永續經營的參考模



式，決定以 500 筆藏品的圖像及詮釋資料作為先導計畫之試行範本。TELDAP 透過計畫內部的一系列座談會議，最後獲得十五組資料集加入 TACP 合作案，內容橫跨藝術、歷史、考古、人類學等學科領域(詳表 4)。

表 4  
參與 TACP 合作案之資料集清單及藏品數量

No.	藏 品	典藏單位	數量
1	明清檔案	中央研究院歷史語言研究所	20 筆
2	佛教石刻造像拓本	中央研究院歷史語言研究所	25 筆
3	考古遺物	中央研究院歷史語言研究所	45 筆
4	甲骨文拓片	中央研究院歷史語言研究所	10 筆
5	漢代石刻畫象拓本	中央研究院歷史語言研究所	40 筆
6	青銅器拓片	中央研究院歷史語言研究所	20 筆
7	善本圖籍	中央研究院歷史語言研究所	20 筆
8	西南少數民族文物	中央研究院歷史語言研究所	20 筆
9	貝類標本	中央研究院生物多樣性研究中心	40 筆
10	古拓碑	國立臺灣大學	20 筆
11	照片	國史館	10 筆
12	書畫	國立故宮博物院	30 筆
13	玉器	國立歷史博物館	18 筆
14	繪畫	華岡博物館	195 筆
15	繪畫、書法、服飾、攝影	TELDAP 第一分項	10 筆
總 合			523 筆

TACP 首先面臨的第一個問題為：應該提供哪些藏品資料給 ARTstor？即藏品詮釋資料應該包含哪些共通性欄位？TACP 首先觀察 ARTstor 實際的詮釋資料案例，分析各個單位諸如美國國會圖書館、大都會博物館、倫敦泰特不列顛藝廊、美國康

涅狄格學院等，發現其提供之藏品詮釋資料欄位從 7 至 14 個欄位多寡不一，其中多數藏品詮釋資料的共通欄位包括：標題(title)、創造者(creator)、權限(rights)、來源(source)四個欄位(詳表 5)。

表 5  
ARTstor 五個單位範例之詮釋資料比對

	Connecticut College	The Metropolitan Museum of Art	J. Paul Getty Museum	Tate Gallery, London	Library of Congress
<b>ID Number</b>	3 cha003	MMA_1977.78			LOC_cai9600008/FP
<b>Title</b>	landscape	Night-Shining White	Rinaldo and the Mirror-Shield	The Lady of Shalott	Interior of a cider mill at Rye
<b>Work Type</b>	scroll	Paintings	Paintings		Drawings and Watercolors
<b>Creator</b>	Cheng Shan-his, Chinese, 1932-	Attributed to Han Gan, Asian; Far East Asian; Chinese, Chinese, active 742-756	Francesco Maffei (Italian (Venetian), about 1605 - 1660)	Waterhouse, John William, English, 1849-1917	Edwin Austin Abbey, 1852-1911, (artist)
<b>Date</b>		Tang dynasty (618-907), 8th century	about 1650 - 1655	1888	1879?
<b>Measurements</b>	Work: 78 x 21 in Image: 19 1/2 x 17 1/2 in	12 1/8 x 13 3/8 in. (30.8 x 34 cm)	Unframed: 30.5 x 34.3 cm (12 x 13 1/2 in.) Framed: 41.6 x 45.4 x 4.4 cm (16 3/8 x 17 7/8 x 1 3/4 in.)		
<b>Material</b>		Handscroll; ink on paper	Oil on copper	o/c	
<b>Location</b>				Tate Gallery, London	
<b>Description</b>		Han Gan, a leading horse painter of the Tang dynasty (618-907), was known for portraying not only the physical likeness of a horse but also its spirit. This painting, the most famous work attributed to the artist, is a portrait of Night-Shining White, a favorite charger of the emperor Xuanzong (712-756). The fairy-tale-like scene, with its burning eye, flaring nostrils, and dancing hooves, evokes Chinese myths about immortal celestial steeds that "sweat blood" and were really dragons in disguise. This unique, procedure-based, scroll fixed by delicate silk threads, is an example of traditional Chinese painting attributed to Tang times.			
<b>Culture Subject</b>	landscape		Italian Knights Genusalemmre liberata Baroque		
<b>Style Period</b>					
<b>Repository</b>	Chu-Gaffis Asian Art Collection, Shein Library, Connecticut College, New London Donated by David Zimmerman '75	The Metropolitan Museum of Art, New York, USA, Purchase, The Dillon Fund Gift, 1977	The J. Paul Getty Museum at the Getty Center 85.PC.321.1		Library of Congress Prints and Photographs Division Washington, DC, USA Cabinet of American Illustration (Library of Congress). (DLC)95849474 Courtesy of The Library of Congress DLC/PP-1994-0044 http://www.loc.gov/rr/print/
<b>related item</b>			<a href="http://www.getty.edu/art/gettyguide/artObjectDetails?artobj=894">http://www.getty.edu/art/gettyguide/artObjectDetails?artobj=894</a>		Published in: "Rye and round there," Harper's magazine, 59:10 (June 1879).
<b>ARTstor Collection</b>	Asian Art Collection (Connecticut College, New London)	The Metropolitan Museum of Art Collection Formerly in The AMICO Library	J. Paul Getty Museum Collection		Library of Congress Collection Formerly in The AMICO Library
<b>Source</b>	Image and original data provided by Connecticut College, New London	Data From: The Metropolitan Museum of Art	Data From: J. Paul Getty Museum	Art History Survey Collection	Data From: The Library of Congress
<b>Rights</b>	Contact Information: Mark Braustein, Visual Resources, Campus Box 5394, Connecticut College, New London, CT 06320-4196, Phone: 860-439-2729, mark.braustein@conncoll.edu Please note that if this image is...	This image was provided by The Metropolitan Museum of Art. Contact information: Photograph and Slide Library, The Metropolitan Museum of Art, 1000 Fifth Avenue, New York, NY 10028, (212) 396-5050 (fax). Please note that if this image is under copyright, you may need to contact one or more copyright owners for	For permission to reproduce images for uses not covered by the Terms and Conditions, please contact Rights and Reproductions, Registrar's Office, The J. Paul Getty Museum <a href="http://www.getty.edu/legal/image">http://www.getty.edu/legal/image</a>	Data from: Digital Library Federation Academic Image Cooperative	This image was provided by The Library of Congress. Contact information: Library of Congress, Photoduplication Service, 101 Independence Avenue, S.E., Washington, D.C. 20540-4570, (202) 707-

其次，TACP 觀察三套國際重要藝術相關的詮釋資料標準及編目規則，分別是藝術作品描述類目 (Categories for the Description of Works of Art, 簡稱 CDWA)、藝術作品描述類目精簡版(簡稱 CDWA Lite)，及描述文物之資料內容標準 (Cataloguing Cultural Objects: A Guide to Describing Cultural

Objects and Their Images, 簡稱 CCO)。對此三套資料結構及資料內容的國際標準，抽取出共通性元素並進行比較分析，發現此三套標準同時將六個欄位列為必填或核心欄位，包含作品類型(work type)、標題(title)、創造者(creator)、日期(date)、材質(material)和典藏單位(repository)(詳圖 3)。

**與國際標準後設資料進行比對，  
發現有6個欄位同時為必填欄位。**

	CDWA	CDWA Lite	CCO
Work Type	Object/Work-Type(物件/作品類型)	1.1 <cdwalite:objectWorkType>	Work Type
Title	Title Text (題名文本)	2.1.1 <cdwalite:title>	Title
Creator	Creator-Description (創作-創作者)	3 <cdwalite:displayCreator>	Creator Display
Date	Creation-Date (創作-時間)	12 <cdwalite:display CreationDate>	Display Date
Material	Materials/Techniques Description (材質/技術描述)	7 <cdwalite:display MaterialsTech>	Material / Technique Display
Repository	Current Repository/Geographic Location (現藏地點/地理位置)	14.1.1 <cdwalite:locationName>	Current Location



圖 3 國際標準比對

最後，TACP 統合上述兩項研究結果，並進一步分析都柏林核心集國際標準(Dublin Core)，以及參與 TACP 合作案之資料集的詮釋資料欄位，建議該合作案的數位藏品提供給 ARTstor 的詮釋資料為：識別碼(identifier)、標題(title)、創造者(creator)、

日期(date)、尺寸(measurements)、作品類型(work type)、描述(description)、材質(material)、主題(subject)、典藏單位(repository)、和權限(rights)等 11 個欄位，而其中 5 個是為必填欄位(標題、日期、作品類型、權利、典藏單位)(詳圖 4)。

**建議TACP與ARTSTOR合作之後設資料**

11個欄位，5個欄位為必填

No.	欄位	屬性	DC	國際標準	ARTstor
1	Identifier 識別碼	選擇	Identifier		
2	Title 標題	必填	Title	Title	Title
3	Creator 創作者	選擇	Creator	Creator	Creator
4	Date 日期	必填	Date	Date	
5	Measurements 尺寸	選擇	Format		
6	Work Type 作品類型	必填	Type	Work Type	
7	Description 描述	選擇	Description		
8	Material 材質	選擇	Material	Material	
9	Subject 主題	選擇	Subject		
10	Repository 典藏單位	必填	Source	Repository	Source
11	Rights 權限	必填	Rights		Rights



圖 4 TACP 與 ARTstor 合作之詮釋資料


2010 年 2 月至 8 月這段期間內，十五組橫跨藝術、歷史、考古、人類學的典藏資源共 523 筆詮釋資料加入 TACP。為減少典藏資源提供者的負擔，TACP 只要求典藏資源提供者提供藏品標題、識別碼及資料庫連結三項資訊(詳圖 5)，TACP 研

究團隊負責從相關資料庫查找出該筆中文詮釋資料，再進行詮釋資料多語化作業。本文將討論 TACP 如何透過四道程序：比對、揀選、翻譯及控制詞彙，建立與 ARTstor 合作的英文詮釋資料。

**由典藏單位提供「選件清單」給TACP團隊**  
●●●(史語所青銅拓片共計20筆)

no.	Title	ID	是否有資料於下列資料庫? 若有「描述」有英文請標明。				選件原因
			聯合目錄	原始典藏庫	知識網	數位島嶼	
1	史獸鼎	187712-079	<input type="radio"/>	<input type="radio"/>			
2	作冊大方鼎	187713-089	<input type="radio"/>	<input type="radio"/>			
3	我方鼎(我麗、禦鼎、禦尊)	187713-108	<input type="radio"/>	<input type="radio"/>			
4	節湯父鼎	187712-080	<input type="radio"/>	<input type="radio"/>			
5	雍鼎(雍作母乙鼎)	188056-002	<input type="radio"/>	<input type="radio"/>			
6	毛公鼎	186789	<input type="radio"/>	<input type="radio"/>			
7	衛鼎	187712-078	<input type="radio"/>	<input type="radio"/>			

1. 選件表單-史語所青銅拓20件.xls



3

圖 5 典藏單位提供之選件清單

(一)比對

參與計畫的十五組數位典藏資源各有其詮釋資料設計，其中部份典藏資源的詮釋資料設計相當仔細且複雜，從各層面描述藏品特色，如考古「遺

物資料庫」詮釋資料除了基本資料，還有銘文、紋飾、形制、工藝分類、表面遺痕等欄位(圖 6)。因此，如何將豐富的詮釋資料正確地比對至簡單的 11 個詮釋資料欄位，是 TACP 很大的挑戰。



圖 6 遺物資料庫的「帶卜辭龜腹甲」(註 4)

TACP 藏品資料主要來源有二，一是呈現藏品豐富內涵的計畫公開網站，另一為收錄所有 TELDAP 計畫藏品目錄的數位典藏與數位學習聯合目錄(註 5)。TACP 研究團隊依據都柏林核心集一對一的原則(註 6)製作藏品比對表，先比對計畫公開網站(即原始網站)與聯合目錄欄位，再依據藏品脈絡揀選適合提供給 ARTstor 的 11 欄位(詳表 6)。TELDAP 計畫常以同一類型的藏品為目標對象(如拓片，通常所有計畫內的藏品都是拓片)，為了區別及分辨藏品，作品類型(work type)常使用較細

緻的下位詞(如炊器(拓片)、水器(拓片)、酒器(拓片)、食器(拓片))，而較少直接採用上位詞(拓片)。然而，當單一計畫藏品資料需呈現於大型且綜合性的資料庫時，如中央研究院歷史語言研究所-青銅器拓片計畫藏品呈現於 ARTstor 平臺時，TACP 建議作品類型(work type)採用上位詞「拓片」，而原先較精細的下位詞則比對至「主題(subject)」欄位。以〈史獸鼎〉為例，TACP 建議提供 ARTstor 之詮釋資料的作品類型(work type)為「拓片」，主題(subject)則是「炊器」(詳表 6)。

表 6  
〈史獸鼎〉提供 ARTstor 之詮釋資料比對

原始網站(器)		原始網站(拓片)		聯合目錄		ARTstor	
原圖名	登錄號	原圖名	登錄號	主要圖名	其他圖名	Title	登錄號
	187712-079		187712-079	登錄號	登錄號		登錄號：187712-079
器號	02778	器號	02778	資料類別	登錄號：187712-079	ID	
				資料類型	類別：炊器	Creator	
				格式	材質：金屬 燒痕： 高寬：67.6 x 45.9	Work Typ	青銅器拓片
		拓號	67.6 x 45.9 cm			Date	
				主編與副編字	銘文內容：尹令史獸立工/于成周·十又一月/癸未，史獸獻工/于尹，成獻工·尹/饒史獸 <sub>二</sub> (禡)，易(禡)冢/鼎一、鼎一，對錫與尹/丕(丕)錫 <sub>二</sub> ，畢(作)父/世	Material	
銘文	尹令史獸立工/于成周·十又一月/癸未，史獸獻工/于尹，成獻工·尹/饒史獸 <sub>二</sub> (禡)，易(禡)冢/鼎一、鼎一，對錫與尹/丕(丕)錫 <sub>二</sub> ，畢(作)父/世	銘文	尹令史獸立工/于成周·十又一月/癸未，史獸獻工/于尹，成獻工·尹/饒史獸 <sub>二</sub> (禡)，易(禡)冢/鼎一、鼎一，對錫與尹/丕(丕)錫 <sub>二</sub> ，畢(作)父/世	日期	西周早期	Measurcm	67.6 x 45.9
原器年代	西周早期	原器年代	西周早期			Subject	炊器
		印記	印記 1 印文：國立中央研究院歷史語言研究所圖書館之印 印記 2 印文：善齋所得彝器 印主：劉體智 印記 3 印文：史語所藏金石拓片之章 印記 4 印文：國院史言	貢獻者	印記印主：中研院史語所 印記印主：劉體智 印記印主：中研院史語所 印記印主：中研院史語所	Description	原器年代：西周早期 印文：國立中央研究院歷史語言研究所圖書館之印(左下1) 印文：善齋所得彝器(左下2) 印文：史語所藏金石拓片之章(右上/右下) 印文：國院史言(左上)
				來源	史語所數位資源整合檢索目錄	Source	聯合目錄 <a href="http://catalog.digitalarchives.tw/dacs5/System/Exhibition/Detail.jsp?OID=8354188">http://catalog.digitalarchives.tw/dacs5/System/Exhibition/Detail.jsp?OID=8354188</a>
原器收藏出土地點	台北-故宮博物院	原器收藏出土地點	台北-故宮博物院	範圍	收藏機構現存地點：台北故宮博物院		
		拓片舊藏	劉體智				
原拓出處	(善齋藏器全形拓本)	原拓出處	(善齋藏器全形拓本)	語言	中文	Related O	02778 原器收藏地：台北-故宮博物院
				管理權	中央研究院歷史語言研究所 ( <a href="http://www.ihp.sinica.edu.tw/">http://www.ihp.sinica.edu.tw/</a> )	Right	中央研究院歷史語言研究所 ( <a href="http://www.ihp.sinica.edu.tw/">http://www.ihp.sinica.edu.tw/</a> )

(二)揀選

第二道程序是揀選，其所面臨最大的挑戰是中文與英文語言上之差異。中國書畫作品上常有創作者或由後人加入的款識或印記，款識文字為沒有標點符號的文言文，即使具備中文書寫能力的現代人也很難正確斷句及判讀。以華岡博物館收藏之〈張

性荃山水〉款識為例(如圖 7 上方的書法)：

問余何事棲碧山笑而不答心自閑桃花流水杳然去別有天地非人間丙辰冬季於養石樓以李白詩意性荃



圖 7 華岡博物館〈張性荃山水〉

此外，部份藏品的款識內容相當長，有時還會出現亂碼(註 7)，如華岡博物館收藏之〈周哲天都峰瀑布〉款識為例：

一生清廉惟黨國 四明佳士育英才 母校  
中國文化大學創辦人張公曉峰夫子百年誕辰誌慶

天都諸峰遙相從 連綿嶂屬無罅縫 山腰白雲  
出衣帶 雲生迭迭山重重 峯內有峰類皺染  
須臾蔚何仍混同 曾雲聚族雨決溜 溪山天水  
齊溟蒙 是時水勢猶未雄 江河欲決翻壑壅  
良久兩足水積厚 瀑布倒瀉天都峰 初疑渴龍  
甫噴薄 抉石投奔聲嗒□ 復疑水激龍拗怒



掉尾下撥百丈洪 更疑群龍互轉斗 移山排谷  
轟圓穹 人言水借風力橫 那知水急翻生風 激  
雷狂電何處起 發作亦在風水中 波浪喧豕草  
木亞 搜攬軒簾心忡忡 潭中老龍又驚寤 緣  
浪瀆湧軒窗東 山根颯拉地軸震 旋恐黃海浮  
虛空 亭午雨止雲戎戎 千條白煉回沖融 憑  
攔心

TACP 經過研究分析後，考量 ARTstor 目標對象為外國人，且詮釋資料將以英文呈現，提出兩項處理原則：

1. 作品款識內容短、辨識度高，則建議翻譯。  
例如華岡博物館所藏之〈金哲夫雪梨秋色〉：

Title	金哲夫雪梨秋色
Description	款識(行書)：金哲夫一九七三

Rights	中國文化大學華岡博物館
--------	-------------

2. 當款識內容太長且出現亂碼，則建議不翻譯。如上述〈張性荃山水〉及〈周哲天都峰瀑布〉；但仍建議可於 ARTstor 平臺呈現藏品款識之中文原文。

### (三)翻譯

翻譯是第三道程序，可以包含四個步驟：譯者初翻、華裔校訂、外籍潤稿、內部總審。上述四個步驟統一於 TELDAP 數位典藏與學習聯合目錄翻譯平台上作業(圖 8)，每位翻譯者都有各自的帳號可以線上登錄，遇到翻譯問題時，舉凡中文詮釋資料的錯字、藏品影像錯誤、翻譯一致性等，翻譯者皆可直接線上標註，方便總編輯及時確認及回應。



圖 8 數位典藏與學習聯合目錄翻譯平台(註 8)



翻譯步驟遇到最大的困難為：藏品詮釋資料具大量藝術專有名詞，這些名詞並無英文對應用法，又屬於少用字彙，查詢不易，導致翻譯速度較慢。以中央研究院歷史語言研究所收藏〈青銅刀柄〉之描述為例：

人形，倒梨狀人面，眉、眼、鼻、齒、唇、耳皆具。……。兩側所見腿部併合方式及裂口不一，革部呈長方形，接刀開口為長橄欖狀，整個器型則為沙漏狀。頂部有一弧形片狀，或許為繫部。

其中「革部」、「繫部」為現代罕見用語，資料難以查詢，對翻譯人員而言較難翻譯。又譬如故宮博物院所藏〈唐人宮樂圖〉之描述為例：

本幅無作者名款。……圖中席面壺門式大案腿下有托泥，形成箱形榻體，四邊有鑿花的銅包角，符合隋唐流行樣式。仕女側坐月牙凳子，是唐代新興家具，長方凳面施錦緞坐墊，四足皆有華麗雕飾，腿間墜以彩穗。此圖飲茶方法屬煮茶法。將調好的茶湯盛入大撇中，再以長杓分茶，盛入茶碗供飲，適

用於人多盛會時。桌面上另置有羽觴，茶酒可同時飲用，並配以茶果。

說明文字中出現「托泥」、「箱形榻體」、「鑿花的銅包角」、「大撇」等專有名詞，翻譯人員參考藏品圖片後仍不易理解，翻譯時只能以模糊的文字處理。

#### (四)控制詞彙

最後一道程序是控制詞彙的應用，控制詞彙來自美國蓋堤研究中心所研發的藝術與建築索引典(The Art & Architecture Thesaurus，簡稱 AAT)及地名索引典(The Getty Thesaurus of Geographic Names，簡稱 TGN)(Getty Research Institute, 2011)(詳圖 9、圖 10)。AAT 運用於作品類型欄位，所採用的控制詞彙有照片(photographs)、油畫(oil paintings)、水彩畫(watercolors)、拓片(rubbings)、書本(books)、善本書(rare books)等；地名索引典只應用於地理資訊是作品脈絡的關鍵資訊之藏品，如照片的拍攝地點、考古物件的發掘地點，則採用地名索引典中的控制詞彙。




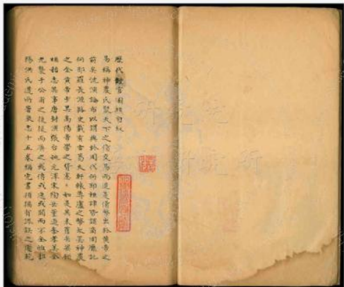
Image Information	
Title	A group photo of chairman of the National Government, Chiang Kai-shek, the American ambassador, Patrick Hurley, and Mao Zedong after a conference. 國民政府主席蔣中正與美國駐華大使赫爾利及毛澤東等人會談後合影
Work Type	photographs 照片
Date	1945/9/25
Measurements	4X5
Description	Acting under orders to facilitate cooperation between the National and Communist parties, the American Ambassador Patrick Jay Hurley poses with Chiang Kai-shek and Mao Zedong for a group photo during a conference in Chongqing prior to his return home to report his progress. 奉命促成國共合作的美國駐華大使赫爾利(Patrick Jay Hurley)於返國進職前，在重慶與進行會談的蔣中正、毛澤東等人合影。
Subject	The President of the National Government   America   Ambassador   Name: Chiang Ching-kuo   Name: Zhang Qun   Location information: Chongqing (China) 國民政府主席   美國   駐華大使   人名資訊: 蔣經國   人名資訊: 張群   地名資訊: 重慶 (中國)
Source	TELDAP Union Catalogs <a href="http://catalog.digitalarchives.tw/?URN=3790552">http://catalog.digitalarchives.tw/?URN=3790552</a> Academia Historica <a href="http://linux201.drnh.gov.tw/textdb/drnhBrowse/advQry.inc.php?_op=?collectionNum:002050102001007">http://linux201.drnh.gov.tw/textdb/drnhBrowse/advQry.inc.php?_op=?collectionNum:002050102001007</a>
ID	典藏號002050102001007

photographs 照片  
AAT ID : 300046300

Chongqing 重慶  
TGN ID: 7024355

圖 9 國史館所照片採用 AAT 及 TGN 控制詞彙的照片

中央研究院歷史語言研究所藏之  
〈歷代鍾官圖經八卷〉



Title	Li dai zhong guan tu jing ba juan 歷代鍾官圖經 八卷
Creator	Chen Xiaolai (Qing dynasty) (清)陳棻孝
Measurements	Collection Length: 2 Volumes Volume Dimensions: 29cm x 29cm 數量: 2冊 外型高廣: 29公分
Work Type	Rare book, Books 善本、圖書
Subject	Rare books -- History -- Metal and stone inscriptions -- Coins Rare books -- Philosophy -- Lists and catalogues -- Artifacts Hand-written copy 善本--史部--金石類--錢幣 善本--子部--譜錄類--器物 鈔本
Source	<a href="http://catalog.digitalarchives.tw/?URN=3205053">http://catalog.digitalarchives.tw/?URN=3205053</a> <a href="http://ndweb.iis.sinica.edu.tw/rarebook/Search/Monograph/Basic/View.jsp?PK_ID=2432">http://ndweb.iis.sinica.edu.tw/rarebook/Search/Monograph/Basic/View.jsp?PK_ID=2432</a>
Rights	The Institute of History and Philology (IHP) of Academia Sinica, Taipei, Taiwan
ID	182771-182772
Description	Version description: hand-written copy Binding: thread-bound 版本敘述: 鈔本 裝訂: 線裝 首有題誰園居士序, 有朱筆校, 沈樹鏞鄭齋收藏印, 群碧樓、鈔本、校本諸印記

rare books 珍本書、善本  
AAT ID : 300266304  
books 書本  
AAT ID : 300028051

圖 10 〈歷代鍾官圖經八卷〉採用 AAT 控制詞彙

### 參、綜合討論

本文在數位圖書館的脈絡下論述多語化的研究方法與框架。藉由執行兩個國際合作研究計畫的研究結果與經驗，說明如何讓全世界不同語言使用者，能更容易檢索與近用數位典藏後中國藝術作品的方法。以下說明兩個多語化計畫所面對的關鍵議題：

#### 一、驗證中英文詞彙語義概念的等同對應程度

分析國立故宮博物院中文控制詞彙，以及 AAT 英文描述語之間的對應關係，先導研究發現中英詞彙之間的等同對應程度，可以區分出六種類型，分別是精準等同、精準等同(cross ref.)、不精

準等同、部分等同(例如：生物類的屬種關係)、不等同(文化特殊性)、不等同(Chen & Chen, 2012)。

本文以其中的不精準等同對應關係為例，並以東西方概念脈絡下的「香爐」(censers)作為說明。美國蓋提研究中心將西方概念脈絡下的「censers」定義為「蓋子有孔的容器，在儀式中用於焚香，特別可見於基督教儀式」，然而，中文概念的香爐定義是「宗教或儀式用的焚香容器，依大小不同可能放置於廟前或是供桌上。」兩相對照下分析出中西方對「香爐」的概念呈現不精準等同關係，因為中文概念裡的香爐可能沒有蓋子，而且並不會用於基督教儀式(圖 11)。在此情況，研究團隊必須為該詞彙增補範圍註描述，以在已有的西方概念之下，也完整呈現東方對此詞彙的核心描述。

A typical censer in the West



Typical censers in Chinese culture

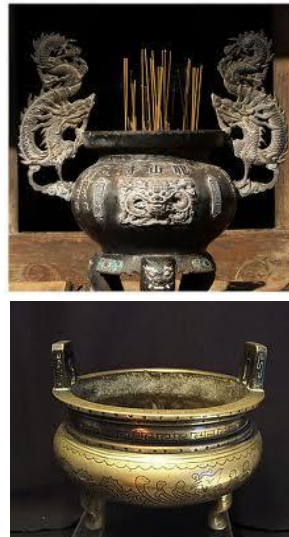


圖 11 東西方概念脈絡下的香爐(censers)

再以部分等同關係為例探討，當詞彙在一個文化脈絡下的概念比另一個文化下的概念涵蓋範圍，較窄或是較廣時，便形成部分等同的對應關

係。本研究發現故宮使用的詞彙概念較 AAT 詞彙精確而深化，因此故宮詞彙經常僅能對應至 AAT 較上位的詞彙。例如：AAT 層級架構裡「幾何紋

飾」的下位詞僅有「小圓點」一個詞彙。相對地，在故宮層級架構裡，「幾何紋飾」的下位詞是「小圓點」，「小圓點」之下還有不止一個下位詞，諸如「乳丁紋」、「穀紋」等。這些「小圓點」之下的下位詞未來需要再進一步檢視，如果分析後可以證實在中國藝術脈絡下是很重要的概念，則本研究建議需要使用創造模組，新增這些詞彙，並放置於目前 AAT「小圓點」層級架構之下。

## 二、詮釋資料品質與成本

TELDAP 以近七個月時間執行與 ARTstor 的合作計畫，透過上述四道程序完成十五組橫跨藝術、歷史、考古、人類學等學科領域的典藏資源，累計共 523 筆詮釋資料。雖完成了有品質的多語化

詮釋資料，然而過程卻花費大量的時間、人力及金錢等成本，特別是在詮釋資料比對及翻譯兩道程序，平均一筆完成四道程序的詮釋資料花費約計 90 美金。

以比對為例，為提高臺灣典藏機構參與本國際合作案之意願，如前所述 TACP 只要求典藏資源提供者提供藏品標題、識別碼及資料庫連結三項資訊，TACP 研究團隊負責從相關資料庫查找出該筆中文詮釋資料，再進行詮釋資料多語化作業。實際執行時遇到兩大困難：其一是不同典藏機構之藏品詮釋資料品質不一，部份藏品需進行中文資料缺字或亂碼確認(圖 12)。

Description	Description (校訂)
說明：此畫為作者於夏天畫此折枝之梔子花來遣興，一枝三花，構圖簡潔，一正一背一苞道盡了梔子花的千姿百態，雖僅工筆淡彩之作，然從畫面上的白描運筆之靈動，花葉設色之簡雅， <b>作者奶O之高不言而喻。</b>	說明：此畫為作者於夏天畫此折枝之梔子花來遣興，一枝三花，構圖簡潔，一正一背一苞道盡了梔子花的千姿百態，雖僅工筆淡彩之作，然從畫面上的白描運筆之靈動，花葉設色之簡雅， <b>作者功力之高不言而喻。</b>
說明：墨分五彩，從單一的墨色中 <b>分出釵h層次</b> ，此幅畫畫面點染，水墨交融，墨韻之變化，層層堆疊，以渲染、皴染手法逐漸轉化，呈現朦朧且變幻多端的形象與氣韻，雲霧或高或低、或曲或折巧妙運轉。龍是神靈之精、四靈之首神，風雲急湧中的龍見首不見尾，看起來活靈活現、栩栩如生，生龍活現，將蓬勃生氣躍然紙上，有神秘之感。	說明：墨分五彩，從單一的墨色中 <b>分出許多層次</b> ，此幅畫畫面點染，水墨交融，墨韻之變化，層層堆疊，以渲染、皴染手法逐漸轉化，呈現朦朧且變幻多端的形象與氣韻，雲霧或高或低、或曲或折巧妙運轉。龍是神靈之精、四靈之首神，風雲急湧中的龍見首不見尾，看起來活靈活現、栩栩如生，生龍活現，將蓬勃生氣躍然紙上，有神秘之感。

圖 12 中文資料缺字或亂碼確認

其次，當藏品資料庫資訊密度過高，特別是資料庫包含兩種形式以上的資源時，如原器及拓片(拓有原器的拓片)，TACP 團隊需要花費較多的人力及時間，才能正確地完成都柏林核心集一對一的原則的比對作業。例如，以青銅器拓片〈史獸鼎〉為例，經過比對發現聯合目錄將「原器年代」誤植為拓片「日期」，因此 TACP 將「原器年代：西周

早期」調整至拓片「描述」欄位(詳表 7)；此外，拓片上的印記資料為研究上重要之參考，因此 TACP 建議以「印文：印文內容(位置)」格式將資料轉入「描述」欄位提供給 ARTstor(詳圖 13)。此部分比對上需要花較多時間去釐清資料庫中不同藏品的層次間關係，才能完成正確的比對。

表 7

〈史獸鼎〉提供 ARTstor 之詮釋資料比對

顏色說明：紅色代表聯合目錄給 ARTstor 的資料，藍色代表本國博物館資料庫給 ARTstor 的資料，綠色代表選件清單給 ARTstor 的資料 選件後設資料表單-史語所青銅拓					
原始網站(拓)			聯合目錄		
原器名	登錄號	拓號	主要題名	其他題名	資料識別
史獸鼎	187712079	02778	史獸鼎	史獸鼎	登錄號：187712079
拓號	02778		資料類型	類別：炊器	
銘文	尹令史獸立工/于成周。十又一月/癸未，史獸獻工/于尹，厥獻工。尹/賞史獸(尊)易(錫)冢/鼎一、鬲一，對揚皇尹/不(丕)顯休，用乍(作)父/厥永寶對卜(尊)彝。	尹令史獸立工/于成周。十又一月/癸未，史獸獻工/于尹，厥獻工。尹/賞史獸(尊)易(錫)冢/鼎一、鬲一，對揚皇尹/不(丕)顯休，用乍(作)父/厥永寶對卜(尊)彝。	主題與關鍵字	銘文釋文內容：尹令史獸立工/于成周。十又一月/癸未，史獸獻工/于尹，厥獻工。尹/賞史獸(尊)易(錫)冢/鼎一、鬲一，對揚皇尹/不(丕)顯休，用乍(作)父/厥永寶對卜(尊)彝。	
原器年代	西周早期	西周早期	日期	中曆：西周早期	
原器收藏	台北-故宮博物院	台北-故宮博物院	來源	史語所數位資源整合檢索目錄	
拓圖			範圍	收藏機構現存地點：台北故宮博物院	
史語所青銅拓_聯合目錄_ARTSTOR.xls			40		

比對欄位：描述	
原器年代	西周早期
印記	印記 1 印文：國立中央研究院歷史語言研究所圖書之記 印主：中研院史語所 位置：左下1 印記 2 印文：善齋所得彝器 印主：劉體智 位置：左下2 印記 3 印文：史語所藏金石拓片之章 印主：中研院史語所 位置：右上/右下 印記 4 印文：國院史言 印主：中研院史語所 位置：左上
Description	原器年代：西周早期 印文：國立中央研究院歷史語言研究所圖書之記(左下1) 印文：善齋所得彝器(左下2) 印文：史語所藏金石拓片之章(右上/右下) 印文：國院史言(左上)

圖 13 〈史獸鼎〉原始網站資料與提供 ARTstor 之「描述」詮釋資料比對



以翻譯為例，藏品「描述」內容經常出現學術性專有名詞，例如缺乏英文對應詞的中國藝術詞彙，需要翻譯者花費較多時間尋找及確認。再加上，有些詞彙過於學術或描述文字太過文言，很難直接翻譯，如何判斷適合的翻譯方式也是一大挑戰。例如藏品「描述」為：「羲之精於章草、隸、八分、飛白、真、行諸體，亦能繪事，後世更尊為書聖」，有兩種翻譯方式：全部翻譯(如 1)或摘要式翻譯(如 2)，也是 TACP 需要面臨的抉擇。

1. Wang Xizhi was skilled in almost all of the calligraphic script types: Zhangcao (one style of cursive scripts), Lishu (clerical script), Bafen, Feibai, Zhengshu, Xingshu (semi-cursive script), etc. He was so influential that later generations elevated him to the supreme position as the “Sage of Calligraphy.”
2. Wang Xizhi was skilled in almost all of the calligraphic script types, and he was already highly regarded in his own time. He was so influential that later generations elevated him to the supreme position as the “Sage of Calligraphy.”

對 TACP 研究團隊而言，所面臨最大的挑戰為：如何在有限的人力、經費及時間內，產出有品質的詮釋資料。檢視先期實驗計畫的四步驟，團隊發現第一個步驟(比對)及第三步驟(翻譯)需要最多的人力、經費及時間，究其原因，前者為典藏單位藏品詮釋資料欄位豐富，如何將豐富的詮釋資料正確地比對至簡單的 11 個詮釋資料欄位，是一大挑戰；後者為藏品描述常出現專業性詞彙或文白交雜的句子，增加翻譯的困難度。為加速多語化詮釋資料過程，以快速地呈現並展示數位典藏，TACP 研究團隊建議未來可採用下述策略：(1)減少詮釋資

料欄位數，只英譯六個(標題、創作者、日期、作品類型、典藏單位、權限)最必要的欄位；(2)先進行合集層次詮釋資料(collection-level metadata)多語化，較屬主題層面的描述先不翻譯，待日後評估再進行多語化；(3)導入 AAT 或任何的 control 詞彙系統於作品類型以維持詮釋資料的一致性。

## 肆、結論

本文以兩個數位典藏多語化國際合作計畫為研究案例，試圖於數位圖書館脈絡下提出資訊組織學的兩個層面之多語化方法架構。古典的資訊組織學包括記述編目與主題編目兩大層面，在本研究以數位圖書館為主要環境的記述編目是以詮釋資料為對象，主題編目則是以索引典為例探討。在詮釋資料方面，本研究個案面對的挑戰是「英文化的範疇」與「英文化的方法」。對於一個已建立百萬件數位典藏的中文詮釋資料，若要逐筆完整地建立英文化詮釋資料，會是一項太浩大而不可及的任務。但為達到多語檢索典藏資源的目標，從 TACP 對於中文詮釋資料的英文化實證研究，本文建議思考三項問題，第一是詮釋資料層次，例如要以合集層次或館藏層次的詮釋資料作為外語化對象，或者兩個層次的詮釋資料皆有需求？若是以館藏層次為主，考量成本效益，是挑選部分藏品只展現代表性的數位資源為主，或者應用控制詞彙諸如質材或類型等屬性讓藏品詮釋資料的英文化詞彙更具一致性，相對也降低成本。若是以合集層次為主，全面轉化為英文的可能性較高。第二是區別詮釋資料的外文化內容，當原來詮釋資料對藏品資源的主題提供詳細而專精的描述文字，從 TACP 研究個案可以觀察到要完整翻譯成另一種語言的效果並不佳，也相對耗費許多資源與成本，因此建議應適當調整數位藏品中英文詮釋資料的詳簡程度及設計。第三是為詮釋資料發展控制詞彙，最基本的範圍類型建議

包含藏品類型、地名、人名及典藏單位名稱。

索引典是控制詞彙的一種類型，可以增進數位圖書館系統的檢索與瀏覽功能與效益。然而，從無到有地發展一套完整的控制詞彙系統需要多年的時間、人力與經費，例如美國蓋提研究中心從 1970 年代便開始發展 AAT，直到目前每年還是需要不斷更新與新增詞彙，以維持索引典的適用性。因此藉由已發展出來的索引典為基礎，進行不同程度的多語化或在地化，是一種常見的發展索引典的研究方法，本文的 AAT-Taiwan 研究個案便是其中的一

項實證研究，提出以翻譯、對應、在地化和創造等四個模組為核心的多語索引典研究框架。此外，基於中西方的語言與文化差異性，以西洋藝術為本所發展的索引典，所呈現的詞彙概念與知識結構的差異性與空缺性，本研究也提出具體解決方法，供作為物質文化或藝術領域之數位圖書館的多語化方法論參考模式，進而有助於日後多語化數位人文的資訊分享。

(收稿日期：2013 年 9 月 1 日)

## 註釋

註 1：本文係根據作者發表於 Chen, S.J., Cheng, C.J. & Chen, H.H. (2011). Methodologies for multilingual information integration in the domain of Chinese art. *IFLA Journal*, 37(4), 296-304. 原稿中文翻譯後，並同時更新與修訂而成。

註 2：「詮釋資料」，是英文 metadata 的中譯，在台灣也譯為「後設資料」。

註 3：<http://www.artstor.org/what-is-artstor/w-html/artstor-overview.shtml>(瀏覽日期：2013/06/04)

註 4：數位典藏與數位學習連合目錄 <http://catalog.digitalarchives.tw/>

註 5：數位典藏與數位學習連合目錄 <http://catalog.digitalarchives.tw/>

註 6：都柏林核心集「一對一的原則」意指都柏林核心集只描述資源的一個表現形式或版本。(Hillmann, 2005)

註 7：亂碼出現原因不一，原因之一為款識書寫的中文字，有些當代已不使用，需要使用造字系統才能呈現。

註 8：<http://www.digital.ntu.edu.tw/teldap/index.php>

## 參考文獻

計畫辦公室(2007)。數位典藏國家型科技計畫第一期(91-95 年度)成果報告(行政院國家科學委員會數位典藏國家型科技計畫技術報告，NSC 95-3113-P-001-003-PO)。台北市：中央研究院。

計畫辦公室(2013)。數位典藏與數位學習國家型科技計畫(97-101 年)總期程成果效益報告(行政院國家科學委員會數位典藏國家型科技計畫技術報告，NSC 101-3113-P-001-003-PO)。台北市：中央研究院。

Borgman, C. L. (1997). Multi-media, multi-cultural, and multi-lingual digital librarians. *D-Lib*, 3(6). Retrieved from <http://www.dlib.org/dlib/june97/06borgman.html>

Chen, H. H. & Chen, S. J. (2008). *Approaches to multilingual access in digital libraries: The case of TELDA*. Retrieved from [http://dublincore.org/groups/languages/dc-2008/Sophy\\_Multilingual\\_Access\\_in\\_Digital\\_Libraries.pdf](http://dublincore.org/groups/languages/dc-2008/Sophy_Multilingual_Access_in_Digital_Libraries.pdf)

Chen, S. J. & Chen, H. H. (2012). Mapping multilingual lexical semantics for knowledge organization systems. *The Electronic Library*, 30(2), 278-294.

- Clough, P. & Eleta, I. (2010). Investigating language skills and field of knowledge on multilingual information access in digital libraries. *International Journal of Digital Library Systems (IJDLs)*, 1(1), 89-103.
- Fox, E. & Marchionini, G. (1998). Toward a worldwide digital library. *Communications of the ACM*, 41 (4), 28-32.
- Getty Research Institute (2011). *Censer in Art & Architecture Thesauurs*. Retrieved from [http://www.getty.edu/vow/AATFullDisplay?find=censer&logic=AND&note=&english=N&prev\\_page=1&subjectid=300198814](http://www.getty.edu/vow/AATFullDisplay?find=censer&logic=AND&note=&english=N&prev_page=1&subjectid=300198814)
- Harpring, P. (2009). *Art & Architecture Thesaurus: Editorial Guidelines*. Retrieved from <http://www.getty.edu/research/tools/vocabularies/guidelines/index.html#aat>
- Hillmann, D. (2005). *Using Duline Core*. Retrieved from <http://dublincore.org/documents/usageguide/>
- Hudon, M. (1997). Multilingual thesaurus construction: Integrating the views of different cultures in one gateway to knowledge and concepts. *Knowledge Organization*, 24 (2), 84-91.
- ISO (1985). *Guidelines for the establishment and development of multilingual thesauri*. ISO 5964.
- ISO (2011a). *Information and documentation - Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval*. ISO/CD 25964-1, ISO/TC 46 / SC 9 ISO 25964 Working Group.
- ISO (2011b). *Information and documentation - Thesauri and interoperability with other vocabularies - Part 2: Interoperability with other vocabularies*. ISO/CD 25964-2, ISO/TC 46 / SC 9 ISO 25964 Working Group.
- Landry, P. (2004). Multilingual subject access: The linking approach of MACS. *Cataloging & Classification Quarterly*, 37(3/4), 177-191.
- Liang, A. and Sini, M. (2006). Mapping AGROVOC and the Chinese agricultural thesaurus: Definitions, tools, procedures. *New Review of Hypermedia and Multimedia*, 12 (1), 51- 62.
- Oard, D. & Diekema, A. (1998). Cross-language information retrieval. *Annual Review of Information Science (ARIST)*, 33.
- Park, J. R. (2007). Cross-lingual name and subject access: Mechanisms and challenges. *Library Resources & Technical Services*, 51(3), 180-189.
- Zeng, M. L. & Chan, L. M. (2004). Trends and issues in establishing interoperability among knowledge organization systems. *Journal of the American Society for Information Science and Technology (JASIST)*, 55(5): 377-395.