

博物館與文化 第4期 頁113~138 (2012年12月)
Journal of Museum & Culture 4 : 113~138 (December, 2012)

數位典藏創意加值之使用者標籤研究¹

陳淑君²

A Study of User Tagging for Creative Added Value of Digital Archives

Shu-Jiun Chen

關鍵詞：使用者標籤、數位典藏、文化創意產業

Keywords: User Tagging, Digital Archives, Cultural & Creative Industries

¹ 本文改寫自筆者參加“International Conference on Asia-Pacific Digital Libraries (ICADL 2011)”國際學術研討會上所宣讀的論文。投稿期間，承蒙匿名審查教授們多方賜教，讓筆者有機會重新反省整篇文章的方法與論述，謹誌謝忱。

本研究為國科會「數位典藏與數位學習橋接計畫」之部分研究成果（計畫編號：NSC100-2631-H-008-002），特此致謝。

² 本文作者現任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.

Email: sophy@sinica.edu.tw

（投稿日期：2012年10月18日。接受刊登日期：101年12月13日）

摘要

本研究採用內容分析法，以 11 位文化創意產業業者對 8 件數位藏品所賦予的 648 個標籤為研究對象，探討數位典藏的商用需求使用者與典藏機構研究員對於藏品索引與檢索時所使用詞彙的落差性，並藉由建立兩者間的語意關係與連結，以提高數位典藏應用加值使用者經驗研究，及改善數位藏品資訊無法完整被商用使用者查詢到的問題。本研究發現越具加值利用可能性的數位藏品，業者給予的標籤數量也越多。業者所有標籤中，大約百分之四十取自典藏機構的學術詞彙，並以四種模式取自典藏機構提供的相關資訊。本研究針對業者自創的商用標籤（60%）與取自學術詞彙的標籤（40%）進行對應分析，發現彼此具有六種語意關係及八種連結屬性。研究顯示目前典藏機構提供的詞彙仍較具領域知識性，商用使用者需要的詞彙相較而言是為更上位概念的詞彙。本研究建議，可將六種語意關係類型及八種連結屬性的研究發現進一步轉化為典藏機構的後設資料與知識組織系統之實踐規範，以橋接商用標籤與學術詞彙之間的落差。

Abstract

With an aim to make digital archives more accessible to cultural and creative industries, this study used, as the subject of research, 648 tags given by 11 practitioners of the cultural and creative industries to 8 digital images in digital archives and analyzed the differences between the terms used by commercial users and those by scholars from archive agencies. We analyzed the self-created commercial tags (60%) and the tags adopted from academic terms (40%). The results showed that the terms provided by the archive agencies were still more likely to be based on domain knowledge. In comparison, the superordinate terms are more likely to be needed by the commercial users. These research findings suggest that six types of semantic relationship and eight types of linked properties could be further transformed into best practices in metadata and

knowledge organization systems for cultural institutions, thus bridging the divide between the commercial tags and academic terms.

前言

使用者標籤(user tagging)是一種俗民分類(folksonomy)，由搜尋資訊的使用者所賦予的標籤(tag)，常與社會性標籤(social tagging)、社會書籤(social bookmarking)、協同式標籤(collaborative tagging)及關鍵字(keyword)等名詞交互使用(Furner, 2007)。國內在此方面的相關研究，對於“tag”的中文翻譯，包括「標記」或「標籤」(卜小蝶，2006；卜小蝶、張淇龍，2009；陳怡蓁，2010)。為行文方便，本文稱之為「標籤」。

使用者標籤源自於 Web 2.0 時代的社群媒體服務，其目的與圖書資訊組織相似，希望有效組織資訊、方便再擷取與利用，但相較於由典藏機構的專業者賦予藏品索引語言的傳統模式，使用者標籤以使用者為中心的典範，提供更貼近使用者的詞彙(卜小蝶，2006；Furner, 2009)。Bar-Ilan (2010: 941)等學者認為此種新興的方法可謂為「群體的智慧」(wisdom of the crowds)，此論點是由 James Surowiecki 於 2004 年所提出，主張在團體中資訊的集合，往往有利於作出比每個成員單獨所做的決定還要好得多的決策(Surowiecki, 2004: 3-22)。此觀點探討三個臭皮匠能勝過一個諸葛亮的哲理，換言之，使用者標籤的精神在於利用群體智慧匯集而成的資訊，將網路價值最大化。

台灣自 2002 年開始由國家科學委員會成立數位典藏國家型科技計畫，包括數百餘個圖書館、博物館、檔案館、學術機構、部會機構，及大學院校的共同參與。至今已產生 400 多萬件數位藏品於網際網路，主題涵蓋人文藝術、歷史、社會科學，至生命科學。該計畫的主要目標是為國家重要而珍貴的典藏品進行數位化工程，以呈現臺灣文化與自然多樣性，並促成典藏內容與技術融入教育、研究、產業、社會及國際化發展(陳淑君、李德財，2010：117-121)。為讓數位藏品能最有效地被使用者檢索，該計畫為每件藏品提供後設資料(metadata)及數位圖檔，並以共通格式的標準放置於數位典藏聯合目錄，讓公眾可以隨時搜尋與利用。後設資料是藏品的背景資訊，包括品名、

類型、時間、創作者、描述等資料，通常是由典藏機構或藏品提供者撰述，著重以學術與專業語彙描述文物特徵，以提供精準的文物內容描述。

然而，當文化創意產業希望利用這些數位藏品加值為商品之際，首先面臨的挑戰便是如何讓業者找到商用目的所需要或感興趣的數位素材。由於不同目的使用者在搜尋數位素材時，會有其特定的情境與需求，及習慣使用的語彙，造成與目前以學術語彙為基礎所建立的藏品描述內容，有部分的差異性（李德財，2010）。因此，本研究以社群媒體的其中一種方式「使用者標籤」，藉其所具備的「群體的智慧」特質，希望及時掌握並探討數位藏品在轉換為文化創意加值用途時，業者標記與文化機構提供的後設資料和學術詞彙之間的落差，及如何以語意關係理論進行連結，作為圖書館，博物館與檔案館知識管理的新方法，以幫助文化創意使用者更有效率地搜尋數位藏品。

使用者標籤與文化資產社群

近年來，許多相關的研究探討使用者標籤如何與傳統文化資產機構，由專業館員或研究人員所賦予館藏的控制詞彙互補，以提供更好的藏品檢索服務(Yi & Chan, 2009; Daly & Ballantyne, 2009; Smith, 2008)。在圖書館方面，Eckert、Hanger 及 Niemann (2009)實驗與比較使用者標籤的品質，該研究發現結合使用者標籤，及人工與自動化索引可獲得最佳品質的文獻註記(annotation)；Lawson (2009)則指出社會標籤能使館員與讀者成為夥伴關係，提供更精進的主題檢索(subject access)服務，該研究量化並比較美國國會圖書館主題標題表，及使用者為亞馬遜網路書店(Amazon.com)與全世界最大的圖書俱樂部 LibraryThing 所進行的三十一個不同主題類別的使用者標籤，找出專業館員與使用者對同樣書籍所賦予的標籤之異同，並藉以論述與評估專業性語彙與使用者用語之間的交叉關係與整合。來自英國普茲茅斯大學創意與文化產業研究所的研究團隊(Alemu, Stevens, Ross & Chandler, 2012)認為在 Web 2.0 時代由社群建構的使用者標籤，仍存在平面、單向與不一致的缺

點，並建議在邁向 Web 3.0 的語意網時代，典藏機構給予藏品的控制詞彙與使用者標籤之間，不應該被看成是兩種截然相反的方式，它們不僅能夠共存，而且應該相互影響彼此，也就是建立混合的後設資料模式。

在博物館方面，最著名也是最大型的實驗計畫是「史帝夫博物館標籤計畫」(Steve.museum tagging project)，該計畫始於 2005 年，旨在協助更多觀眾使用數位藏品。近年數位博物館逐漸盛行之際，許多博物館與美術館發現線上參觀者在數位藏品搜尋方面遇到困難，輸入關鍵字卻找不到所要的藝術品。此問題主要源自於藝術史學家所撰寫的博物館藏品語彙和一般觀眾使用語言上有許大的落差。因此該計畫希望以社會標籤和大眾分類法，幫助觀眾更有效率地搜尋藝術品，並縮小觀眾和博物館之間語意上的鴻溝，另一方面，博物館則可藉以增加跨文化和跨語言的觀點，使更多不同類型的觀眾使用和欣賞線上藏品(Trant, 2006)。為了探討社會標籤在博物館中的角色與效益，史帝夫團隊使用實驗法來測試不同情形下的使用者反應，並蒐集有關使用者的量化和質化資訊。研究結果顯示社會標籤對於博物館說明文撰寫是一個相當具有效益且獨特的工具，並可以加強線上蒐藏的使用率。該計畫提出兩項具體的研究發現：(一)將使用者標籤與博物館標籤進行比對，顯示 86% 的貢獻標籤沒有出現在原來的博物館後設資料資料庫之中，換言之，超過八成的使用者標籤與博物館說明文完全沒有任何重覆，顯示一般觀眾對於藝術品的語言認知和博物館有很大的落差；(二)將標籤與其他博物館使用的專業術語，如藝術與建築索引典(Art & Architecture Thesaurus, AAT)和藝術家人名權威檔(Union List of Artist Names, ULAN)進行比對，結果顯示 70.2% 的使用者標籤與 AAT 詞彙相符，及 36% 的使用者標籤與 ULAN 詞彙相符(Trant, 2009)。

台灣的數位典藏計畫，每件藏品的後設資料多以學術性語彙為主，以國立故宮博物院在數位典藏聯合目錄的一件數位藏品「北宋 汝窯 青瓷水仙盆」為例，後設資料的內容如下（北宋汝窯青瓷水仙盆，1101-1150）：

後設資料元素	資料內容
資料類型	陶瓷器
描述	橢圓形盆，侈口、深壁，平底凸出窄邊稜，四雲頭形足；周壁胎薄，底足略厚。通體 滿佈天青釉，極勻潤。裹足支燒，底部有六個細支釘痕，略見米黃胎色。全器釉面純潔無紋片。
日期	1101 A.D.-1150 A.D.
關聯	《千禧年宋代文物大展》 《大觀—北宋汝窯特展》 《故宮瓷器錄 第一輯 宋元》(上編)

從此實例可以觀察到幾項典藏機構提供後設資料的特徵：(一) 講求文獻保證原理，以上例的「關聯」元素可以看出此後設資料的文字內容，主要參考三份出版品；(二) 著重以學術與專業術語描述文物特徵，例如在「描述」元素，使用「...侈口、深壁，平底凸出窄邊稜，四雲頭形足；周壁胎薄，底足略厚。通體 滿佈天青釉，極勻潤。裹足支燒，底部有六個細支釘痕...」的描述語言，多是學者之間論述器物的專業術語。因此，在數位典藏的產業化研發過程，基於以學術為導向的藏品描述，如何被業者更容易接近使用是其中一項重要議題。

綜合上述的相關研究，可以觀察到使用者標籤的實驗對象多為網路上的民眾，並未區分特定領域的群眾。其中的一項重要探討議題，是使用者標籤與控制詞彙（也就是學術與專業語彙）之間的比較研究，而研究結果也顯示使用者標籤確實有其顯著的貢獻，但隨著語意網的環境逐漸成熟，使用者標籤與學術詞彙之間的整合與連結將是 Web 3.0 時代的重要研究方向。

研究方法與理論基礎

本研究先以檔案類數位藏品為對象，首先觀察與分析文化創意業者提供給數位藏品的使用標籤，辨識學術詞彙與商用詞彙之間的落差程度；其次分

析業者利用學術性詞彙的模式；最後分析商用性詞彙與學術性詞彙的語意關係及連結屬性的類型。

一、研究對象

本研究採用量性與質性兼具的內容分析法，以文化創意產業業者對數位檔案所賦予的標籤(tag)為研究對象，探討數位典藏的商用需求者與典藏機構的專業者，所使用詞彙的差異性。本研究首先以典藏於中央研究院歷史語言研究所的「內閣大庫檔案」為對象，此檔案原藏於清宮內閣大庫，內含各色明清時期（1368—1911）文書（中研院史語所歷史文物陳列館，2006）。基於典藏機構的後設資料，多以專業詞彙描述藏品的背景與特徵，不易被文創業者解讀，因此本研究請該典藏單位提供其認為具商用潛能的數位藏品，並為每筆藏品提供約 200 字具故事性的藏品描述及關鍵字，再讓業者依此知識為基礎，重新賦予以加值利用為角度的標籤。由於時間的限制，僅先請典藏單位提供 8 筆數位藏品資料作為本研究的基礎。

其次，選取文化創意相關業者，包括教育出版及數位學習、生活應用及文化商品、數位娛樂及影音頻道及視覺藝術及數位藝術等 4 個領域別，總計 11 家業者，分別為惠茂印刷設計、達志影像、傳揚、創意連結、開物國科技文化、佳世達科技、子生文創、呂豪文工業設計、彼得潘藝術、遠百企業及中華平面設計協會（詳表 1）。

本研究請上述業者逐一檢視 8 筆內閣大庫檔案的數位圖檔及藏品描述後，從商用的角度給予每筆藏品兩項資訊，分別是「加值利用可能性程度」及「關鍵字」（詳表 2）。

表 1 文化創意業者之領域別與屬性

業者公司名稱	業者領域類別	業者屬性
惠茂印刷設計有限公司	教育出版及數位學習	印刷裝訂、設計包裝
達志有限公司	教育出版及數位學習	圖庫經紀代理
傅揚實業有限公司	生活應用及文化商品	燈飾製造
創意連結有限公司	生活應用及文化商品	品牌代理、通路經營、授權經紀、商品開發、肖像開發
開物國科技文化事業有限公司	生活應用及文化商品	文創商品開發
佳世達科技股份有限公司	數位娛樂與影音頻道	電腦及消費性電子製造業
子生文創事業有限公司	視覺設計與數位藝術	衣飾設計
呂豪文工業設計有限公司	視覺設計與數位藝術	工業設計
彼得潘藝術有限公司	視覺設計與數位藝術	文創商品開發與通路
遠百企業股份有限公司	視覺設計與數位藝術	百貨通路
中華平面設計協會	視覺設計與數位藝術	識別形象設計、商業包裝設計、品牌策略與形象規劃

(資料來源：本研究整理)

表 2 內閣大庫檔案之數位圖像與使用者標籤量

藏品編號	品名	業者標籤的數量
G001	長江圖殘卷局部	70
G002	山西省墩堡圖殘本局部	69
G003	包袱式樣(盛京)	107
G004	包袱式樣(皇史宬)	98
G005	乾隆己卯科廣西鄉試題名錄封面	81
G006	夏至天象方位圖	75
G007	雲龍邊框局部(嘉順皇后仙逝敕諭)	75
G008	雲龍邊框局部 (出使英國欽差大臣劉瑞芬坐名敕書)	73
合計	648	

(資料來源：本研究整理)

二、分析方法

本研究採用內容分析法，進行業者對每個藏品所賦予關鍵字的歸類與屬性分析，包括量性與質性研究技巧。為瞭解學術與商用使用者對藏品關鍵字的異同程度，本研究先蒐集 11 位業者對 8 件數位藏品總計 648 個關鍵字，接著分別將每件藏品關鍵字與典藏機構提供的描述文字逐一核對，標示、編碼並計算出 (A) 同時出現在典藏機構描述文字的關鍵字，及 (B) 只有出現在商用使用者標註的關鍵字等兩類關鍵字。接著，剔除每筆藏品重複出現的關鍵字，並將這些關鍵字的編碼聚合在一起，再重新計算出上述 (A) 與 (B) 類關鍵字數量與分布。

在 (A) 類方面，本研究以逐筆典藏品為分析單位，就關鍵字與描述文字之間的關係，逐一檢視與歸納類型，以瞭解業者如何直接利用典藏機構有用的詞彙作為其標籤的資訊行為。在 (B) 類方面，本研究同樣以逐筆典藏品為分析單位，將關鍵字進行與藏品之間關係的語意關係及連結屬性分析與歸類。

所謂語意關係是指概念間或意義間的關係，如「鼎」是一種「中國禮器」，這說明了兩個概念或意義之間的階層關係；「毛公鼎」是周宣王年間所鑄造的一個「鼎」，此則顯示此兩者之間具有實例與概念關係。在資訊檢索中，語意關係的基本功能有助於增加數位藏品的檢全率(recall ratio)與檢準率(precision ratio)。根據 Milstead(2001: 65)的研究發現，建議應該發展出更豐富且更具階層組織的關係，因為資訊搜尋者需要的是一種類似地圖的圖像，可以告知他們所存在與活動的世界。此論點正好可以檢視典藏機構所給予的學術詞彙是否能夠滿足商用使用者對於藏品的檢索效益。本文以丹麥皇家圖書資訊學院 Birger Hjørland 教授歸納的重要語意關係為理論基礎(Hjørland, 2007: 403-405)，進行實證研究。Hjørland 舉出的十七種語意關係，包括：一、主動關係(Active relation)；二、反異關係(Antonymy)；三、聯想關係(Associative relation)；四、因果關係(Causal relation)；五、同音異義關係(Homonymy)；六、下位詞關係(Hyponymous relation)；七、類例關係(Instance-of relation)；

八、位置關係(Locative relation)；九、整部關係(Meronymy)；十、被動關係(Passive relation)；十一、聚合關係(Paradigmatic relation)；十二、多義詞(Polysemy)；十三、所屬關係(Possessive)；十四、相關詞彙(Related term)；十五、同義詞(Synonymy)；十六、時間關係(Temporal relation)；十七、方式關係(Troponymy)。本研究在業者標籤與學詞彙的語意關係分析與歸類方面，包括如下三個程序：

- (1) 研究者徵求四位實作者進行比對實作，並為實作者舉行實作前的訓練與練習，以讓其了解如何進行此項任務，及比對的步驟與判斷技巧。
- (2) 實作者被分為兩組，同時進行只出現在商用標籤的 354 個詞彙之比對分析，並詳細記錄不確定或無法判斷的比對個案。
- (3) 研究者根據實作者完成的初步比對結果與記錄，逐一檢視與確認。尤其是對於兩組實作者之間的比對有歧異、或無法確認者，定期與實作者進行深度討論，以建立更明確及一致的比對規則。

研究結果與討論

本研究以典藏機構所提供之內閣大庫檔案共 8 件數位藏品為基礎，每件藏品包括數位圖檔、品名及藏品描述等三項資訊，再由 11 位業者檢視上述資訊後，從文化創意的應用角度分別為每件藏品提供標籤，共獲得 648 個標籤作為分析樣本。表 3 為業者加值利用可能性分布情形，整體而言，業者認為超過三分之二的數位藏品具有加值利用的可能性（51%）及極有可能性（17%），大約三分之一無加值利用的可能（32%）。從此數據可以觀察到典藏機構認為具商用潛能的 8 件藏品，經過業者檢視數位圖檔及藏品描述後，只有三分之二被認為具有商業加值的可能性。換言之，典藏機構與業者之間對於數位藏品的商用可能性，存在大約三分之一的認知落差。根據表 3 分布情況，可以觀察出藏品（G003）是被最多業者認為「極有可能」加值利用，及最少業者認為「不可能」加值利用者。將此進一步與表 2 的業者標籤數量交叉比較可發現，該件藏品也是本研究範圍所有藏品中，收集到業者最多數

量的標籤。

表3 業者加值利用可能性之分布

藏品編號	極有可能業者數量	有可能業者數量	無可能業者數量
G001	1	5	5
G002	0	6	5
G003	5	4	2
G004	5	3	3
G005	2	7	2
G006	0	6	5
G007	1	7	3
G008	1	7	3
合計(百分比)	15 (17%)	45 (51%)	28 (32%)
全部總計	88 (100%)		

(資料來源：本研究整理)

一、學術與商用標籤之異同分析

業者的所有標籤，在本研究中進一步區分為取自學術(A類)或自創商用(B類)兩類標籤，如表4所示分布，可知不論是否剔除重複出現的標籤，總體而言，以B類由業者對數位藏品新增的標籤佔較多，計354個(佔55%)，倘使去除重複的標籤則佔更多，計329個(佔62%)。換言之，典藏機構提供的說明文字，大約百分之四十為業者所採用作為標籤。由此或可推知目前典藏機構所提供以學術用途為主的說明文字，可能尚有六成無法立即滿足欲對數位藏品進行商用的使用者之需求。究竟此六成由業者新增的標籤與四成取自典藏機構的標籤之間，是否有任何連結性與語意關係？針對此部分的議題，本研究將於本文後面進一步探討與分析。

此外，表4顯示其中藏品(G006)標籤分布狀況明顯異於整體的平均

分布甚多，其 A 類佔 72%、B 類佔 28%，表示業者大多採用典藏機構的詞彙作為標籤，本身再新增的標籤很少。進一步觀察該筆藏品的內容，發現典藏機構提供的說明文字長度是其他七筆藏品平均內容的 2.8 倍，而其圖像是唯一以簡易的黑白呈現。將此藏品（G006）與業者加值利用可能性（詳表 3）交叉比較可發現，其為三件被最多業者列為較無可能加值利用的其中一件。與此相對的是藏品（G007）的標籤分布狀況剛好與藏品（G006）相反，其 A 類佔 28%、B 類佔 72%，表示業者大多是自己再新增標籤，較少採用典藏機構的詞彙作為標籤。進一步觀察該筆藏品的內容，發現典藏機構提供的說明文字長度是所有八筆藏品中最少內容者。此現象似乎顯示（一）典藏機構的藏品描述越詳細，業者採用為標籤的比例也越高；（二）數位圖像本身呈現的豐寡訊息可能也是影響業者給予新標籤的數量，及判斷是否可能進一步加值利用的基礎。

表 4 學術與商用標籤之異同分布

藏品編號	標籤數量總計		去除重複的標籤數量	
	A* (百分比)	B** (百分比)	A* (百分比)	B** (百分比)
G001	32 (46%)	38 (54%)	23 (40%)	34 (60%)
G002	33 (48%)	36 (52%)	253 (41%)	33 (59%)
G003	54 (50%)	53 (50%)	31 (38%)	50 (62%)
G004	45 (46%)	53 (54%)	31 (39%)	49 (61%)
G005	33 (41%)	48 (59%)	26 (37%)	45 (63%)
G006	54 (72%)	21 (28%)	32 (62%)	20 (38%)
G007	21 (28%)	54 (72%)	15 (23%)	50 (77%)
G008	22 (30%)	51 (70%)	18 (27%)	48 (73%)
合計	294 (45%)	354 (55%)	199 (38%)	329 (62%)
(百分比)	648 (100%)		528 (100%)	

註：1.*表示 A 類標籤，是業者取自典藏機構提供的品名或藏品描述文字之標籤。
2.**表示 B 類標籤，是業者對數位藏品新增的標籤，並未出現於典藏機構提供何資訊。

（資料來源：本研究整理）

二、業者利用典藏機構詞彙為標籤之模式分析

將業者直接採用典藏機構的 294 個所有詞彙進一步觀察與分析，發現無論是來自藏品品名或藏品描述文字，具有「完全照錄」、「切割概念」、「取用上位概念」及「組合部分概念」等 4 種標籤的採用模式，以下分別就此 4 種標籤採用模式進行討論。

(一) 完全照錄

業者完全採用典藏機構提供的品名或藏品描述文字，作為其標籤。如：「長江圖殘卷局部」(TB0122, TB0143)該標籤完全採自藏品(G001)的品名。

(二) 切割概念

業者先將典藏機構提供的品名或藏品描述文字切割數個概念，作為其標籤，而非完全照錄地視為一個標籤。如：藏品(G002)的品名「山西省墩堡圖殘本局部」，業者(T0243)以切割方式採用此品名作為標籤，包括「山西」、「墩」、「墩堡」、「山西省」。

(三) 取用上位概念

業者檢視典藏機構提供的描述文字後，只採用其中較上層概念的詞彙為標籤。如：藏品(G003)「包袱式樣」在本研究樣本的一段如下描述文字：

「此幅為清代所繪盛京應用織金坐龍黃緞地黃綢裏夾包袱式樣，原件尺寸：縱 112 公分，橫 112 公分。觀看時應順時針右轉 45 度，此樣精繪正龍紋與皇帝龍袍同式，皆正面向前呈坐姿，具有威嚴。左、右及上方各有代表富貴之牡丹花紋，象徵祥瑞之兆的雲紋點綴其間，顯得相當生動靈活，下角正中裝飾著波濤翻騰的浪花及點點飛沫，別有一番趣味。」

其中業者(T0344)只取其中「龍袍」與「祥瑞」兩個詞彙作為商用標籤，另一業者(T0342)則取其中「皇帝龍袍」與「牡丹花紋」兩個詞彙作

為商用標籤。此類模式顯示業者以典藏機構提供的詞彙為標籤的參考，並選取其中相較於較為上層概念的詞彙，而非其中的學術專業詞彙。

(四) 組合部分概念

業者對於典藏機構提供的描述文字並未完全照錄，而是取部分概念再組合成一個標籤。例如：藏品 (G004)「包袱式樣」在本研究樣本的一段如下描述文字：

「此幅為墨筆所繪皇史宬應用織金龍黃緞地黃綢裏夾包袱式樣...所繪為五爪雙遊龍，龍與蟒之分在爪，龍有五爪，乃皇帝、皇后、皇太子、親王、皇室成員所用紋飾；三爪、四爪為蟒，供一定品級之官員使用。...」

業者 (T0422) 借用其中「皇室成員」及「紋飾」，並組合成「皇室成員紋飾」為概念的標籤。甚至，此類模式也出現借用典藏機構的部分詞彙，再與業者本身的詞彙組合而成一個新的標籤。例如：藏品 (G001)「長江圖殘卷局部」在本研究樣本的一段如下描述文字：

「此幅長江圖殘卷...大約完成於明永樂六年至宣德五年(1408-1430 A.D.)，左自湖廣岳州府臨湘縣境內鴨欄磯起，右至興國州上山磯止...」

業者借取上述的「明永樂」詞彙，再自己給予「古地圖」詞彙，組合成「明永樂古地圖」為概念的標籤。

三、商用標籤與學術詞彙暨藏品之關連分析

前述分析結果顯示業者所有標籤中，有六成是由業者新增的標籤，四成取自典藏機構的詞彙。因此，本研究嘗試透過對每件藏品的此兩類詞彙對應分析探討這兩者之間，所能夠建立的語意關係及連結屬性。首先就標籤之間的語意關係進行檢驗，以瞭解商用標籤與學術詞彙的橋接方式；其次分析商

用標籤與學術詞彙暨藏品的連結屬性，藉此瞭解看似鬆散的商用標籤與嚴謹的學術詞彙及藏品圖像之間具有哪些可互相接軌的重要屬性。

(一) 商用標籤與學術詞彙之語意關係分析

本研究針對商用標籤與學術詞彙之間的語意關係進行辨識，結果如表 5 所示，共分析出 6 種語意關係的類型，分別為「屬種關係」、「類例關係」、「同義關係」、「整部關係」、「位置關係」及「聯想關係」，以下逐一討論各種語意關係的現象及如何改善商用標籤與學術詞彙之間的語意落差。

「屬種關係」(S001) 意指商用標籤與學術詞彙之間呈現一種上、下位詞的關係，在本研究的樣本顯示出商用標籤對於藏品的詞彙需求較學術詞彙更為上位，如：商用標籤使用「中國紋飾」(TR0345) 表達學術詞彙使用的「牡丹花紋」(TB0412) 概念。此現象意謂典藏機構可能因掌握更多此藏品的知識內容，而使用較為精確而特定的詞彙。若是可以藉由知識組織系統為學術詞彙自動建立其上位詞，並與藏品連結與索引，將可進一步改善基於屬種關係的語義落差問題。

「類例關係」(S002) 是指商用標籤與學術詞彙之間呈現概念與此概念的個別實例之語意落差，如：商用標籤使用普通名詞「傳教士」(TR0623) 表達學術詞彙使用的專有名稱「湯若望」(TB0612)。根據藝術史學者 Panofsky 的圖像學理論，此現象意謂典藏機構對於「是什麼」(of-ness) 作品的詞彙已經超過前圖像學的「描述」層次，到達表現上的分析之「識別」層次(Panofsky, 1972: 5-17)。然而，商用業者對於數位藏品的實際經驗或熟悉性所客觀認知到的仍是在基本的一般性詞彙。

「同義關係」(S003) 表示商用標籤所指的意義與學術詞彙相同，但雙方使用不同的詞彙表達等同的概念，例：「清代」(TB0312) 與「清朝」(TR0323) 互為同義詞關係。同義詞是一個概念在廣泛脈絡中，意義相同或幾乎相同的詞彙(Harpring, 2010: 3)。以搜尋為目的時，可以將一組視為相等的詞彙組織起來成為同義詞環(synonym rings)，與檢索引擎一起使用，幫助擴充檢索式。

使用者只要使用同義詞環中的任何一個詞，就可以找到這類的資源（曾蕾，2006）。根據本研究的歸納，可以知悉蒐集與建立同義詞並建立同義詞環，將有助於改善數位藏品被商用業者檢索時的檢全率(recall ratio)。

「整部關係」(S004)係指商用標籤與學術詞彙之間呈現概念的整體與部分的關係。此類關係的部分詞指的是某事物更小的組成部分、物質或成員

表 5 商用標籤與學術詞彙之語意關係類型

類型 編號	語意關係類型	意涵	研究例證	
			學術詞彙(A)	商用標籤(B)
S001	屬種關係	上、下位詞的關係，說明一個物件列屬於哪個分類(如：A 是 B 的一種)。	牡丹花紋 (TB0412)	中國紋飾 (TR0345)
S002	實例關係	指一般概念與此概念的個別例子之間存在的語意關係(如：A 是 B 的一個實例)。	湯 若 望 (TB0612)	傳 教 士 (TR0623)
S003	同義關係	A 所指的意義與 B 相同；A 與 B 相等。	清 代 (TB0312)	清 朝 (TR0323)
S004	整部關係	整體與部分的關係(A 是 B 的一部分)。部分詞指的是某事物的組成部分、物質或成員。	親王、皇后、 皇太子、皇帝 (TB0412)	皇 族 (TR0423)
S005	位置關係	以一個概念說明一個物件在另一個概念指定下的位置(如：A 位於 B)。	湖廣岳州府 臨 湘 縣 (TB0112)	中 國 (TR0112)
S006	聯想關係	以心理的方式來定義的關係，多半為不明確的關係。在本研究可以包括：衍生、創意及應用聯想等三類。	龍(TB0823)	降 龍 十 八 掌、亢龍有 悔、龍戰於 野、潛龍勿用 (TR0823)

(資料來源：本研究整理)

，如：商用標籤使用「皇族」(TR0423)此一整體概念，表達學術詞彙所使用的「親王」、「親王」、「皇太子」、「皇帝」(TB0412)等相較下更小的組成分子。

「位置關係」(S005)在此是商用標籤與學術詞彙之間形成一種地理位置的語意關係，如：商用標籤使用「中國」(TR0112)表達學術詞彙使用的「湖廣岳州府臨湘縣」(TB0112)。從此例證也可以觀察到，商用標籤的詞彙是較以淺層的一般性概念或專有名稱為主；相較之下，典藏機構提供較多更細節的概念或專有名稱，但也忽略在專指度高的概念之外，連同提供其較上層的概念，而這些似乎正是商用業者對於數位藏品圖像所產生的「原型」概念。

「聯想關係」(S006)是以心理方式來定義的關係，多半為不明確的關係。本研究意指商用標籤與學術詞彙之間存在聯想性，如：商用業者以「降龍十八掌」、「亢龍有悔」、「龍戰於野」及「潛龍勿用」(TR0823)等武俠小說中與學術詞彙使用的「龍」(TB0823)相關的各種武學招式。根據本研究歸納兩者之間的語意關係，可以包括衍生、創意及應用聯想等三種次類的聯想關係。此類的語意落差，可能需藉由使用者標籤系統的機制，以定期蒐集相關詞彙，並開放讓更多其他使用者讀取這些詞彙，再形成更多的聯想概念及各種可能的文化創意連結。

(二) 商用標籤與學術詞彙暨藏品之連結屬性分析

本研究欲瞭解未出現在典藏機構提供的詞彙範圍，由業者對每件藏品所標籤的詞彙之屬性。換言之，此類型詞彙是存在典藏機構與業者之間的落差，若能辨識出這些詞彙與典藏機構已提供的學術詞彙、或與藏品之間的關係屬性，當能進一步提供縮短學術（或專業）與商用索引詞彙的差異問題。本研究針對只有出現在商用使用者標籤的關鍵字總計 354 個詞彙，進行逐一檢視、分析與歸類後，如表 6 所示，共得 8 種類型的標籤與藏品之連結屬性，分別是「時間」、「空間」、「類型」、「材質」、「物理特質」、「學科」、「人物」，

及「商用功能」。依 Layne 對於圖像的屬性分類，本研究結果可以對應至經歷屬性、主題屬性及關係屬性等三大類(Layne, 1994: 583-585)。茲以 Layne 的圖像屬性為基礎，就標籤與數位藏品之間的連結屬性，描述如下：

表 6 商用標籤與學術詞彙暨藏品之連結屬性類型

類型編號	連結屬性類型	對應的 Layne 圖像屬性分類	意涵	研究例證	
				學術詞彙(A)	商用標籤(B)
R001	時間	經歷	藏品產出的中國朝代	明永樂六年 (TB0112)	明朝 (TR0123)
R002	空間	經歷	藏品產出的國家或地方	盛京 (TB0312)	中國 (TR0112)
R003	類型	主題	藏品本身的種類屬性	圖 (TB0242)、 乾隆己卯科廣西鄉試題名錄 (TB0512)	地圖 (TR0212)、 中國古書 (TR0523)
R004	材質	主題	藏品製成所使用的材料	曆書 (TB0645)、 地黃綢 (TB0312)	紙 (TR0623)、 絲綢 (TR0423)
R005	物理特質	主題	藏品的外觀、形狀、色彩、圖案、花紋與裝飾	立姿正龍 (TB0512)、 皇史宬應用織金龍黃緞 (TB0412)	龍形 (TR0545)、 金色 (TR0423)
R006	學科	主題	所屬的學門或領域	天文 (TB0612)	中國天文科技 (TB0623)
R007	人物	主題	由其職業或從事的活動定義出來的個體或群體	光緒 (TB0823)	天子 (TR0823)
R008	商用功能	關係	從業者觀點，認為未來可以加值應用的產品項目	包袱 (TB0423)、 圖 (TB0142)	四方巾 (TR0423)、 筆記本 (TR0422)、 絲巾 (TR0422)、 屏風 (TR0422)、 藏寶圖 (TR01232)

(資料來源：本研究整理)

(一) 經歷屬性

根據 Layne 的論點，這個類別的屬性與圖像的由來有關，本研究的結果顯示業者著重於圖像的「誕生」資訊需求，包括「時間」(R001)及「空間」(R002)類別。「時間」(R001)是指藏品產出的時間，本研究的樣本顯示，業者常使用朝代作為標籤的對象，如「明朝」(TR0123)；「空間」(R002)則是藏品產出的國家或地區，如「中國」(TR0112)。因此典藏機構可以檢視目前的數位藏品後設資料，是否已規劃「日期」及「地理空間」元素，以滿足使用者的資訊搜尋需求。

(二) 主題屬性

數位藏品的主題是指描述、識別或解釋作品所含的知識內容，本研究歸納出文化創意業者在此類最常使用的屬性包括：「類型」(R003)、「材質」(R004)、「物理特質」(R005)、「學科」(R006)及「人物」(R007)等四個面向。「類型」(R003)是藏品本身的種類屬性，例如商用業者使用圖(TB0242)、地圖(TR0212)、名錄(TB0512)、古書(TR0523)等詞彙。「材質」(R004)是藏品製成所使用的材料，例如商用業者使用紙(TR0623)、絲綢(TR0423)等詞彙。「物理特質」(R005)則是藏品的外觀、形狀、色彩、圖案、花紋與裝飾等特徵，例如商用業者使用龍形(TR0545)、金色(TR0423)等詞彙表達其對該數位藏品外觀的概念或認知。「學科」(R006)意指所屬的學門或領域，例如業者使用中國天文科技(TB0623)詞彙表達其對該數位藏品所屬領域的註記。「人物」(R007)是指由其職業或從事的活動定義出來的個體或群體，本研究觀察到業者對於數位藏品所呈現出來與人物有關的概念，有相當程度的興趣，例如業者使用天子(TR0823)等詞彙註記其對某個作品的重要概念。從此部分的探討，可以知悉典藏機構在為數位藏品規劃後設資料時，對於上述四個面向的主題屬性，需提供相對應的元素及控制詞彙，才能夠讓業者如期地搜尋到其所需要的藏品。

(三) 關係屬性

數位藏品可能與其他影像、文學作品或可以應用的商品相關，因此為藏品指出這些關係的存在及本質是讓使用者可以搜尋或進一步延伸探索的重要資訊。本研究所指的關係屬性，是從業者觀點認為未來可以加值應用的產品項目，例如業者使用四方巾（TR0423）詞彙表達其對典藏機構的包袱（TB0423）藏品的可能的加值應用產品。除此，業者也使用諸如：四方巾（TR0423）、筆記本（TR0422）、絲巾（TR0422）、屏風（TR0422）、藏寶圖（TR01232）等詞彙。

結論與建議

本文從文化創意產業業者的角度探討數位典藏的商用需求者與典藏機構的專業者，對數位典藏所賦予的標籤之差異性。研究中除針對內閣大庫檔案被賦予的學術與商用詞彙的落差性加以分析，並對業者利用學術性詞彙的模式、商用性詞彙與學術性詞彙的語意關係，及連結屬性等進行實證研究。綜合本研究結果，提出具體結論與相關建議如下：

一、數位典藏機構的描述性詞彙可以滿足四成的商用標籤

本研究發現，典藏機構提供的數位檔案經業者檢視後，具有超過三分之二的加值利用可能性。其中，越具加值利用可能性的數位藏品，業者給予的標籤數量也越多。本研究同時發現，業者所有標籤中，大約百分之四十取自典藏機構的詞彙，主要採用自藏品品名或藏品描述文字，具有「完全照錄」、「切割概念」、「取用上位概念」及「組合部分概念」等四種不同模式。本研究結果代表業者對於典藏單位的後設資料描述之關鍵詞截取模式，建議典藏機構在後設資料描述之際，也可以參照依此四種模式，擷取關鍵詞作為數位藏品的索引詞，以改善業者資訊檢索的檢準率。

二、商用標籤與學術詞彙之間的落差，可藉由六種語意關係予以連結

本研究結果顯示，業者所有標籤中，大約百分之六十是由業者新增的標籤。究竟這樣的數據是太多還是太少？代表的意義是什麼？此結果相較於「史帝夫博物館標籤計畫」的研究結果，百分之八十六的使用者標籤並未出現在原來博物館的後設資料，似乎本研究的業者採用更多典藏機構對藏品的描述詞彙。不過，需指出的是這兩個研究的基礎並不相同，「史帝夫博物館標籤計畫」在使用者貢獻標籤時，並未事先提供後設資料，此意謂該研究數據是使用者的認知未受影響下的結果。而本研究，則是基於典藏素材作為商用加值時，業者認為缺乏故事性可能是造成利用的阻礙因素之一（何文雄，2010），因此先請典藏機構提供藏品的故事性描述，如此業者必然會被典藏機構的描述所影響。在此，百分之四十到底有多少是受其影響，這部分的問題並未在本研究設計中進一步區分出來。本研究僅關注於當典藏機構已提供藏品的故事性描述之情況下，仍有百分之六十無法滿足業者需求的詞彙，該如何解決。

針對自創商用標籤（60%）與取自學術詞的標籤（40%），本研究在進行上述兩類標籤的對應分析，發現六種語意關係類型，分別為「屬種關係」、「實例關係」、「同義關係」、「整部關係」、「位置關係」及「聯想關係」。本研究提出具體改進策略，是建立商用與學術標籤之間的語意連結，藉由兩者可能的連結關係類型的歸納，以促進之間的互動。因此本研究建議，可將此六種語意關係類型的研究發現進一步轉化為典藏機構的知識組織系統之實踐規範，以橋接商用標籤與學術詞彙之間的落差。此外，相較於 Hjørland 所歸納的十七種語意關係，本研究的分析結果僅出現六種，是商用與學術詞彙僅存在此六種關係，或因受本研究的分析樣本數量與類型限制而無法呈現其它語意關係，是未來需再進一步釐清的研究，例如，將來可再以圖書館及博物館的數位藏品為基礎，並增加樣本數量，以彌補本研究在取樣的限制下，可能造成所得結果普遍性降低的問題。

三、商用標籤主要取自數位藏品的八種屬性

本研究結果發現只出現在商用使用者的自創標籤與典藏機構已提供的學術詞彙、或與藏品之間，可以歸類出八種類型的標籤與藏品之連結屬性，分別是「時間」、「空間」、「類型」、「材質」、「物理特質」、「學科」、「人物」，及「商用功能」。換言之，商用使用者在搜尋數位檔案時，會以上述八種屬性為基礎嘗試找出需要的數位藏品，雖然整體而言，典藏機構目前已大致提供這些屬性的詞彙，但研究顯示目前這些詞彙仍較具領域知識性，商用使用者需要的詞彙相較而言，是為更上位的詞彙。本研究建議，如同上述可將此八種連結屬性的研究發現進一步轉化為典藏機構的後設資料之實踐規範，以橋接商用標籤與學術詞彙之間的落差。

參考文獻

- 卜小蝶，2006。淺談社會性標籤之意涵與應用。在 Web 2.0 與圖書館研討會論文集，頁：1-17。新北市：淡江大學。取自 http://www.lib.tku.edu.tw/libintro/pub/web20&lib_semina/social_tag_ft.pdf（瀏覽日期：2011年2月16日）
- 卜小蝶、張淇龍，2009。社會性書籤網站之使用者與標籤特性初探，圖書資訊學研究，4（1），1-26。
- 北宋汝窯青瓷水仙盆（1101—1150）。數位典藏與數位學習聯合目錄。取自 <http://catalog.digitalarchives.tw/item/00/0c/bf/cb.html>（瀏覽日期：2012年8月28日）
- 李德財，2010。數位典藏與數位學習橋接計畫—總計畫。行政院國家科學技術發展基金補助計畫績效評估報告（NSC 97-3114-P-001-002-Y）。臺北市：中央研究院資訊科學研究所。
- 何文雄，2010。數位典藏與數位學習橋接計畫—創意整合加值典範計畫。行政院國家科學技術發展基金補助計畫績效評估報告（NSC 97-3114-P-307-006-Y）。臺北市：財團法人資訊工業策進會。
- 陳怡蓁，2010。層面分類結構應用於社會性標記之研究：以圖書作品為例。國立臺灣大學圖書資訊學研究所碩士論文，臺北市。
- 陳淑君、李德財，2010。國家型數位典藏計畫之回顧、現況與展望：以台灣為例。第八次中文文獻資源共建共享合作會議圖書館中文資源與數位典藏學術研討會論文集，頁 113-140。台北市：國家圖書館。
- 中研院史語所歷史文物陳列館，2006。內閣大庫檔案（三）。取自 http://www.ihp.sinica.edu.tw/~museum/tw/artifacts_from_ethnic.php?class_plan=135（瀏覽日期：2011年2月5日）
- 曾蕾，2006。用於標引、瀏覽、檢索的知識組織系統。在數位圖書館發展國際研討會，頁 1-80。台北市：國立臺灣大學。取自 [http://moodle.lips.tw/~tcasist/activities/2006/ISDDL2006/Knowledge%20Organization%20Systems%20\(KOS\)%20for%20Indexing,%20Browsing,%20and%20Retrieval%20.pdf](http://moodle.lips.tw/~tcasist/activities/2006/ISDDL2006/Knowledge%20Organization%20Systems%20(KOS)%20for%20Indexing,%20Browsing,%20and%20Retrieval%20.pdf)（瀏覽日期：2010年11月19日）
- Alemu, G., Stevens, B. Ross, P. & Chandler, J., 2012. Linked data for libraries: benefits of a conceptual shift from library-specific record structures to RDF-based data

- models. *In: Proceedings of World Library and Information Congress: 78th IFLA General Conference and Council*, pp. 1-17. Helsinki, Finland: IFLA.
- Bar-Ilan, J., Zhitomirsky-Geffet, M., Miller, Y. & Shoham, S., 2010. The effects of background information and social interaction on image tagging. *Journal of the American Society for Information Science and Technology*, 61(5): 940-951.
- Daly, E., Ballantyne, N., 2009. Ensuring the discoverability of digital images for social work education: an online “tagging” survey to test controlled vocabularies. *Weology*, 6(2), Retrieved January 20, 2011, from <http://www.webology.org/2009/v6n2/a69.html>
- Eckert, K., Hanger, C., Niemann, C., 2009. Tagging and automation: challenges and opportunities for academic libraries. *Library hi Tech*, 27(4): 557-569.
- Furner, J., 2009. Folksonomy. *In: Bates. M. J. (Ed), 2009. Encyclopedia of library and information sciences*(3rd Ed), Retrieved January 18, 2011, from <http://polaris.gseis.ucla.edu/jfurner/papers/Folksonomy3.pdf>
- Furner, J., 2007. User tagging of library resources: toward a framework for system evaluation. *In: Proceedings of World Library and Information Congress: 73th IFLA General Conference and Council*, pp.1-10. Durban, South Africa: IFLA. Retrieved December 20, 2010, from <http://archive.ifla.org/IV/ifla73/papers/157-Furner-en.pdf>
- Harpring, P., 2010. *Art & Architecture Thesaurus: Editorial Guidelines*. Los Angeles: J. Paul Getty Trust. Retrieved January 15, 2011, from http://www.getty.edu/research/conducting_research/vocabularies/editorial_guidelines.html#aat
- Hjørland, B., 2007. Semantics and knowledge organization. *Annual Review of Information Science and Technology*, 41: 367-405.
- Lawson, K., 2009. Mining social tagging data for enhanced subject access for readers and researchers. *The Journal of Academic librarianship*, 35(6): 574-582.
- Layne, S. S., 1994. Some issues in the indexing of images. *Journal of the American Society for Information Science*, 45(8): 583-588.
- Milstead, J. L., 2001. Standards for relationships between subject indexing terms. *In: Bean, C. A. & Green, R. (Eds.), 2001. Relationships in the Organization of Knowledge*, pp. 53-66. Dordrecht: Kluwer Academic Publishers.

- Panofsky, E., 1972. *Studies in Iconology-Humanistic Themes in the Art of the Renaissance*. Boulder: Westview Press.
- Smith, G., 2008. Tagging: emerging trends. *Bulletin of the American Society for Information Science and Technology*, 34(6): 14-17.
- Surowiecki, J., 2004. *The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business*. New York: Doubleday.
- Trant, J., 2006. Exploring the potential for social tagging and folksonomy in art museums: proof of concept. *New Review of Hypermedia and Multimedia*, 12(1): 83-105.
- Trant, J., 2009. *Tagging, Folksonomy and Art Museums; A report of the steve.museum research*. Retrieved July 10, 2012, from <http://conference.archimuse.com/files/trantSteveResearchReport2008.pdf>
- Yi, K., Chan, L. M., 2009. Linking folksonomy to library of congress subject headings: an exploratory study. *Journal of documentation*, 65(6): 872-900.