

數位典藏內容建置平台發展

曾慕曦 劉吉軒

國立政治大學資訊科學系

台北市指南路二段 64 號

g9111@jsliu@cs.nccu.edu.tw

摘要

數位典藏內容建置一直以專業典藏機構為主，但是有鑒於相關單位所能涵蓋的文化產物畢竟在質與量均甚為有限，許多民間文化資源隨著時間的流逝很容易就消失了，所以我們希望讓一般人對生活週遭值得典藏的文化及事物也能進行紀錄與保存。同時我們認為，典藏對象的數位化與多媒體化，必須結合跨領域的知識與訓練，如：主題領域、數位媒體、資料管理、人機介面、軟體工程等。結合以上的目標，我們開發了數位典藏內容建置平台提供給使用者建立典藏文化的管道與方法。

關鍵字

主題式多媒體歸檔，數位內容保存與歸檔，團隊合作，ontology

1. 緒論

隨著資訊科技的快速發展，電腦世界中的資料流通已經深植在使用者的日常生活中，二十一世紀人們使用電腦的頻率大大提升，拜電腦科技進步之賜，用多媒體來傳輸檔案已經是不可阻擋的趨勢，由於多媒體的變化與多樣性，讓使用者在運作上能夠完整表達出欲呈現的意涵，也讓多媒體資料提供更豐富的紀錄空間，因此網路上的數位檔案數量如同炸彈開花一樣成爆炸性的成長。絕大多數使用者皆已慢慢熟悉如何將實體物件數位化，並且由於寬頻網路的普及，資訊所能到達的範圍實在是無遠弗屆，因此不論身在何處只要連接上網路，不需要花費大量的成本與等待的時間，便能夠輕鬆取得遠在數百公里之外的數位內容；導因於

此，我們發現如何建置一個數位典藏內容建置平台讓數位化媒體素材儲存、歸檔以及再應用是很重要的課題。

長久以來，世界各國的相關組織一直大力推行且投入大量資源的數位典藏工作僅限於高價值的歷史文物，如：故宮博物院，國外的 American Memory、British Museum、Louerea Museum 等等，相較於此，一般民間文物缺乏有效的儲存和管理，隨著時間的流逝很可能會消失不見，若是一段時間後要加以紀錄與觀察其難度相當的高，我們認為，若是能將一般人的文物都加以保存，並且有系統的規劃欲保存的規格和架構，以此建置而成的數位典藏內容資料庫必定兼具豐富的深度和廣度。

政府機構所能投入的人力與資源畢竟有限，而且能夠典藏的涵蓋面也有限，弱勢族群與民間文化傳承將隨著時間而消失，原因就在於相關機構無法將資源投入到此領域之中，問題就卡在人力與資源的不足，但是經過多次相關數位典藏實作的經驗我們發現，具有相關系統開發能力的人才大大的存在於學校、社團、家庭等，雖然這些人的能力只有業餘的水準，但是經過適當的整合與協調，往往都能夠做出職業般的系統作品。

在本篇論文我們定義了完成數位典藏工作的工作流程，並且把角色與人物的團隊合作關係定義出來，然後再參考工廠生產線的想法，把工作的過程以物件傳遞的方式運作，以此為建立成 ontology 去驅動數位典藏內容建置平台的運作，透過此 ontology，我們實作了 HAM(Heritage Archiving Mediator)，實際地讓

使用者能透過 HAM 典藏文化遺產，並且學習如何進行數位典藏的工作。

經由 HAM 的運作，我們完成了顧正秋藝術網站、政治大學傳播學院影像資料庫與政治大學圖書館與檔案人物網站，透過三個專案的印證，我們得到豐富的典藏結果。

以下的論文我將於第二章描述相關數位典藏工作的文獻探討，討論目前數位典藏工作的進程，並且提出需要解決的問題。第三章則提出資源需求的分析，用來結合專業技能與知識領域。第四章提出 HAM 的建置與系統架構。第五章則是利用 HAM 開發出來的實例網站。第六章是結論。

2. 文獻探討

Witten 對於數位典藏做了以下的定義[2]，"A collection of digital objects, including text, video, and audio, along with methods for access and retrieval, and for selection, organization and maintenance of the collection."。數位典藏目的在於把圖書、文化、器物、遺產等，經由數位化的過程，轉換成可儲存的檔案，並且將檔案完整描述存入資料庫，爾後透過全文影像以及 metadata 的整合，融入圖文影音等資料，提供給使用者瀏覽、檢索、整合資訊與增值應用。

數位典藏的工作可區分為圖書館與博物館，前者的研究重點在於圖書、期刊雜誌與論文，目標是讓使用者快速的檢索取得相關資訊[5]；而數位博物館的研究重點在於文物的典藏[6]，隨著歷史的前行，許多珍貴文物不斷的產生，為了妥善保存這些珍貴文物，如果不經由數位化的處理，恐怕很難辦到。數位典藏在政府大力支持下，相關研究已經相當多年，其中包括故宮博物院、自然科學博物館、中央研究院和國家圖書館等知名機構均投入大量的人力、物力。

康乃爾大學於 1998 年發表的 NCSTRL 數位圖書館[4]，不僅保存了實體圖書館的資料，並且透過 indexing 和包裝，整合成 repository 的

集合，經由 user interface 讓使用者快速的檢索數位圖書館。

香港中華大學的 iVIEW 數位圖書館[1]於 2002 年對於 digital library 的多語系有初步的琢磨，舉凡 indexing、searching、retrieval，都能透過不同的語言來 query 出相關的結果，其做法是抽取 video 和 audio 的 feature 結合成單一獨立的 modal information，然後用 natural multilingual text 來 query 這個 modal，以得到理想的 data，雖然目前成果僅限於英文和中文，但仍不斷努力中。

Schilit 於 1998 年對於數位圖書館的應用有更多的發展[3]，由於一般的數位圖書館的重點都放在檢索的部分，而且都需要電腦才能利用相關增值功能，本研究把數位圖書館的應用延伸到平板電腦、手機等等，可以做到 active reading、free form ink annotation、fluid movement among document activities、physical mobility。

個人認為數位典藏除了上述的發展以外，仍有值得深入研究和改進的地方。目前數位典藏的涵蓋面太小，能夠納入的資訊量與範圍有限，同時我們進行數位典藏的時候，缺少跨領域的結合，許多典藏工作必須要有足夠的技術支援，其前提就是質與量並重的人力資源。此外我們還認為，數位典藏的工作流程沒有完善的規劃，往往只注重在特定工作，如果可以引入工作流程的觀念，把數位典藏內容的工作規劃清楚，就能讓相關研究者很清楚的執行，也能大大的提高效率。

3. 資源需求分析

要完成數位典藏內容建置系統必須考量到各個領域人才的參與和合作，清楚定義各項工作的階段性任務，再者如何將人才與工作之間透過角色與階段性物件作媒介，結合系統的功能，讓數位典藏內容建置系統過程運作流暢，成功將數位內容產出成多媒體資料庫和使用者介面。

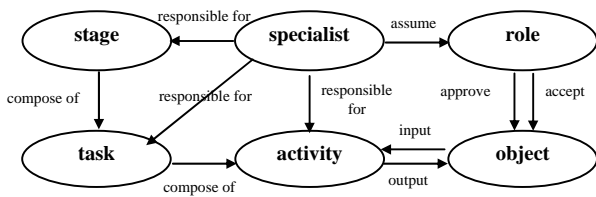


Figure. 1 Schematic diagram of domain ontology

現今數位典藏的建置缺少的就是普及化與平民化，由於數位典藏內容建置一直以專業典藏機構為主，因此所謂的團隊合作已經變成很自然的事情，反而相關研究人員沒有清楚的描述其分工的必要性，導致後來的研究者沒有一套依循的機制，對於數位典藏的工作只能自行摸索。

一般人大多不了解生活週遭的許多文化及事物也能進行紀錄與保存，因為沒有類似的經驗與典藏的風氣，若是要引領數位典藏的風氣就是要提供一套方法與有效的平台供使用者使用，如何協助使用者了解數位典藏的重要性，並且加以教導其典藏自身的文化與事物是很重要的事情。

首先我們必須定義數位典藏的 ontology，藉由 ontology 已經整合的目標與功能架構，驅動系統的運作，讓系統發揮相關領域人才的知識整合(specialist)、工作流程的規劃(stage、task、activity)、角色扮演(role)、物件輔助(object)等功能。(Figure. 1)

3.1 相關人員的合作

依據近年來數位典藏的工作經驗，我們認為”人”是進行相關工作最基本的元素，在工作的過程中，每個人的背景與知識也都不一樣，所以定義所需的人才並且整合能力是必須努力的方向。以下我們將相關領域的人才細分成以下五種專業人員(specialist)：

- **Subject Domain Specialist**

專案的起始者與領導者，對整體系統規劃有一系列的思考藍圖，主導相關典藏系統的開發有豐富的經驗，而且在各大領域上的知識均有涉獵。

- **Digital Media Specialist**

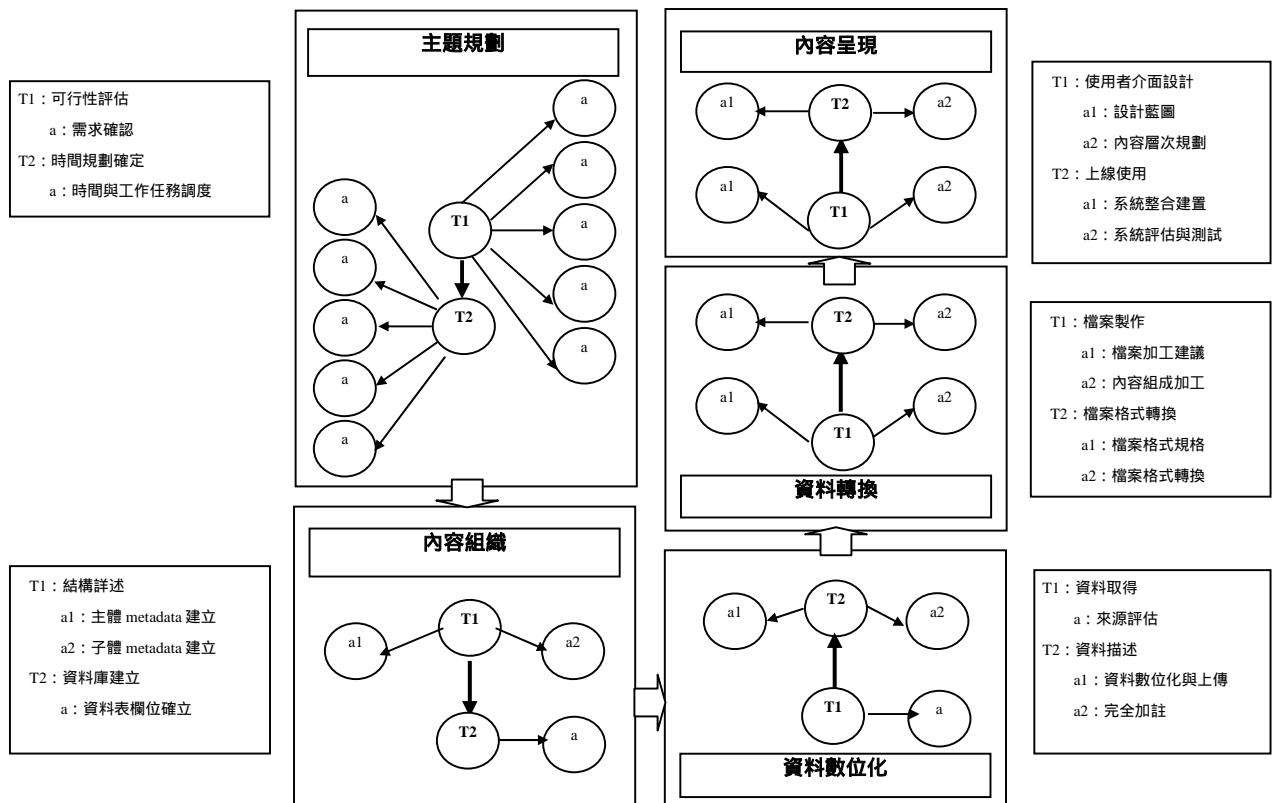


Figure.2 Hierarchical Decomposition of Digital Archiving Workflow

長期接觸與製作數位媒體的相關產品，對於媒體的特性與操作相當熟稔，了解每一份數位媒體內容背後所需蘊含的 metaphor，在數位媒體典藏實作的過程中，點出數位媒體內容最精要的部分，並且以適當的手法呈現。

- **Data Management Specialist**

蒐集相關資料的人員，有著豐富的檔案管理經驗，了解表現一個特定主題所需素材的質與量；同時在檔案的管理和描述上，熟知需要內涵哪些相關知識，並且把相關的知識文字化、視覺化到數位典藏內容建置系統。

- **Graphic Interface Specialist**

設計數位媒體內容前端使用者介面的人員。此專業人員往往有著相當多實作的經驗，除了本身有很強的美工設計技術以外，都有初階網頁設計的能力，這對於如何將數位媒體內容前端介面，不失真的交給軟體工程專業人員做資料庫與網頁技術的連結有很大的幫助。

- **Software Engineering Specialist**

包裝數位媒體內容資料庫與數位媒體內容前端介面成為數位典藏內容建置系統的人員，撰寫網頁程式，包括前後端的系統建置管理，整合各領域專家的需求，實作出最後的系統；同時在系統實作的過程中，提供各個專家於電腦上的技術支援。

3.2 工作流程的建立

將數位典藏導入工作流程的概念，就是要把每個典藏工作步驟都定義清楚。我們界定每一個 stage 都有其獨立性，前一個 stage 結束後才進入下一個 stage，同時我們還依據每個工作階段的特性，將其都更細分到 task 和 activity。Figure. 2 是工作模組的流程圖。

Stage 1 : 主題規劃 (Subject planning)

工作開始我們必須針對未來實作時所需的各項軟硬體設備與時間工作分配做規劃。我們將此 stage 細分為可行性評估和時間規劃確定兩個 task。

- **Task 1 : 可行性評估 (Feasibility evaluation)**

在此任務我們主要希望得到各個專家未來實做時需要用到的軟硬體設備，藉此了解工作的需要。在此設定一個 activity 需求確認。藉由系統內部已經設計好的軟硬體設備進行勾選。

- **Task2 : 時間規劃 (Plan and schedule development)**

我們希望讓每個 specialist 提出自身的工作時間，以及所須負擔的工作進度。在此設定一個 activity 時間與工作任務調度。預先評估系統各階段所需的人力與資源，依據每個成員的專業知識加以分配工作，並且規劃階段性任務成員的責任與權重。

Stage 2 : 內容組織 (Content organization)

在資料蒐集之前，要先訂定未來要蒐集的素材分類，再根據每個分類所需要的相關資訊做細部設計。此 stage 細分為結構詳述和資料庫建立兩個 task。

- **Task1 : 結構詳述 (Structure specification)**

將未來數位媒體素材的規格建立出來，蒐集素材時只要依循著特定的規範，就不需要擔心事後需要補全任何的資訊。本 task 可分為 subject domain specialist 需要負責的主體 metadata 建立和 data management specialist 需要負責的子體 metadata 建立兩個 activity。

- **Task2 : 資料庫建立 (Database development)**

依循上一個 task 設計好的 metadata 建立資料庫，本 task 僅分為一個由 software engineering specialist 負責的資料表欄位建立的 activity。

Stage 3 : 資料數位化 (Material digitalization)

相關人員開始著手蒐集資料的動作，無論是以什麼方式取得，都必須遵循既定的資料格式；此外每一份素材還需要有加註的動作，讓素材進一步轉化文件。我們將此 stage 細分為資料取得和資料描述兩個 task。

- **Task1 : 資料取得 (Material acquisition)**

多方面採集資料並且核定適當資料，界定所需的資料範圍，包含人事時地物，以便往後的蒐集、加註與數位化。在此設定一個 data management specialist 負責的來源評估一個 activity。

- **Task2：資料描述 (Material annotation)**

每一份資料都有其資訊與知識蘊藏其中，因此做註解與描述有其必要性，如此一來，完整的數位媒體素材就順利被文件化。本 task 區分為由 digital media specialist 負責的資料數位化與上傳以及由 data management specialist 負責的完全加註兩個 activity。

Stage 4：資料轉換 (Data transformation)

每一個檔案要正式上線呈現在使用者面前時，還必須先經過最佳化的動作，例如：檔案的串流化、影像影片高低階轉換。本 stage 分為檔案製作和檔案格式轉換兩個 task。

- **Task1：檔案製作 (Data authoring)**

數位文件之間有相當的關聯，若是能夠加以整合，可以讓文件增值。本 task 分為由 subject domain specialist 負責的檔案加工建議和 digital media specialist 負責的內容組成加工兩個 activity。

- **Task2：檔案格式轉換 (Data format transformation)**

數位化檔案距離真正上線後的檔案仍有一段距離，如檔案的備份、檔案品質的提升、檔案傳輸的加速等。在此分為 data management specialist 負責的檔案格式規格與 digital media specialist 負責的檔案格式轉換兩個 activity。

Stage5：內容呈現 (Content presentation)

使用者介面依據相關的資料作調整與規劃，包括視覺設計、操作便利性以及使用者喜好，並且將介面與資料庫做結合。系統完成後，還須經由多方面的測試。在此 stage 分為使用者介面設計和上線使用兩個 task。

- **Task1：使用者介面設計 (Interface design)**

資料庫與使用者的溝通管道是人機介面，把數位文件忠實呈現出來到使用者面前是重要的工作。以下分為 graphic interface specialist 負責的設計藍圖與內容層次規劃兩個 activity。

- **Task2：上線使用 (System online)**

系統在使用者介面設計完成以後，就是與資料庫做連結，在系統完成後，需要開放給外部使用者修正。本 task 分為由 software engineering specialist 負責的系統整合建置與全體人員都需負責的系統評估與測試。

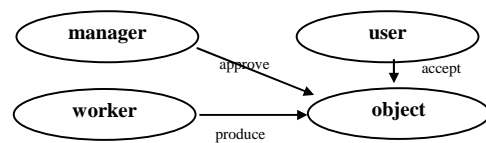


Figure.3 Relationships between Role and Object

3.3 工作過程中的角色

一部電影會有很多角色參與其中，完成數位典藏工作也不例外。在相關工作階段中，我們定義了三種角色。Figure. 3 是角色在工作流程中需要發揮的功能。

Manager

工作任務的領導者，除了參與一部分的工作，對於全部工作的進度和熟悉度必須最高，必須適時給予工作人員意見。在每一份工作任務結束時，審核評估是否同意(approve)接受階段性完成物件進入下一個工作任務。

User

每個任務進行之前，都會有 specialist 決定是否 accept 上一個工作任務傳遞下來的 object，因為工作的傳承有一致性，每個環節所產生的結果都必須加以檢查才能繼續下去。

Worker

每個工作任務的特性與需求均不相同，所需的人才也不同，因應這樣的需求，分配任務給相關領域的工作人員，此人員必須負責將自己的任務完成，並且產出階段性物件。

3.4 階段性物件的產出

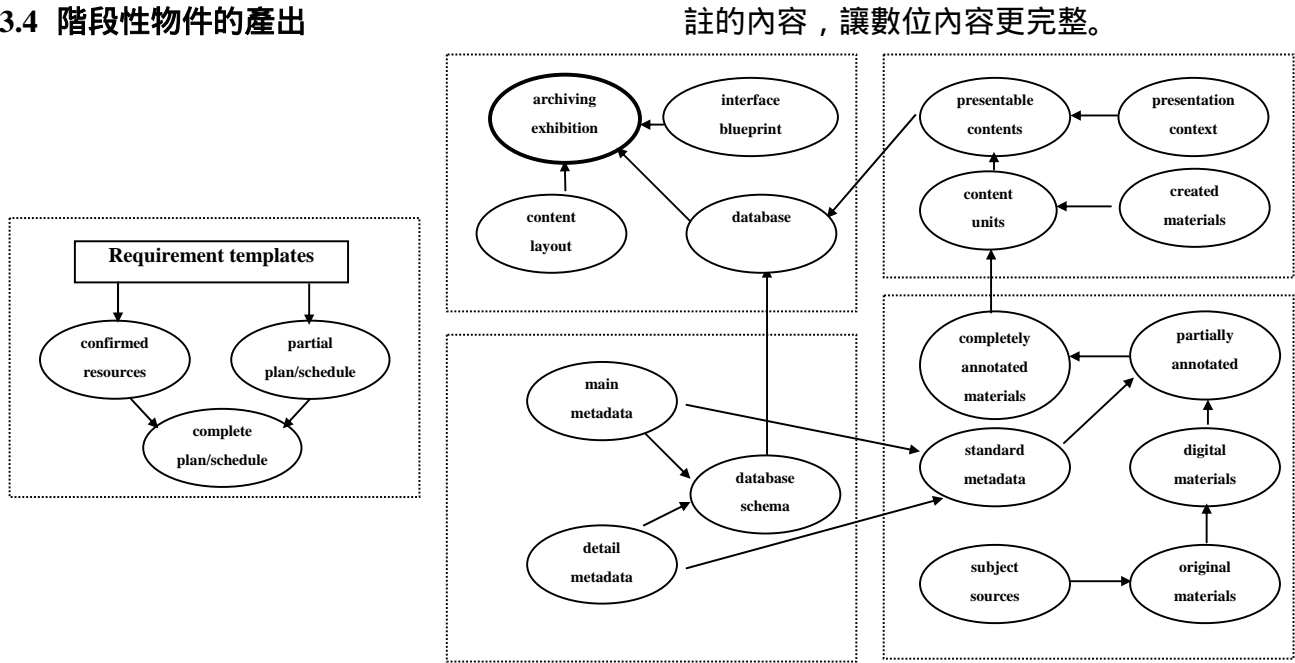


Figure. 4 Object Transformation Process

驅使工作流程的運作順暢，是基於物件的前後傳遞，把物件之間的從屬關係界定清楚，對於系統的完成有莫大的幫助。Figure. 4 是物件傳遞過程的示意圖。

經過需求確認和時間與工作任務調度得到每個成員需要的資源與時間規劃。再從數個主要後設資料群中，由使用者選取符合製作主題的後設資料，透過後設資料建立資料表欄位。之後使用者必須開始評估如何取得數位素材，得到素材來源，然後再透過管道取得相關原始素材，把所有的素材數位化，此時透過之前的後設資料，實做出介面提供給使用者上傳到系統之中，不過由於上傳的使用者對數位素材的認知有限，此時的數位素材只能算是部分加註的數位內容，必須透過其他的使用者加以強化加

再來的目標是把完整數位內容做加值的處理與轉換，相關的數位內容可以透過整合，或是使用者的創意與想法加工結合，產製出新的數位內容。由於數位內容最後必須呈現在網際網路，所以必須要有呈現的規格，讓數位內容轉換成適合網際網路的呈現格式。最後則是將數位內容整合入資料表，進行資料庫的整合與正規化，依據不同數位內容的風格，先實做出使用者前端的藍圖，然後把藍圖結合實際數位內容，數位內容建置系統便宣告完成。

4. Heritage Archiving Mediator

基於上一章節的資源分析，我們可以把各塊資源加以整合，並且以此為 ontology 開始發展系統平台，我們把它稱之為 HAM (Heritage Archiving Mediator)。Figure. 5 是 HAM 的架構圖，我們把整套系統分為實體層次與概念層次兩大主軸；實體層次包含、使用者介面設計與多媒體資料庫系統；概念層次則包含以數個為具體功能層面，串聯階段性物件於平台的工作流程之中，然後透過 network flow 的連結，把實體層次和概念層次結合在一起。

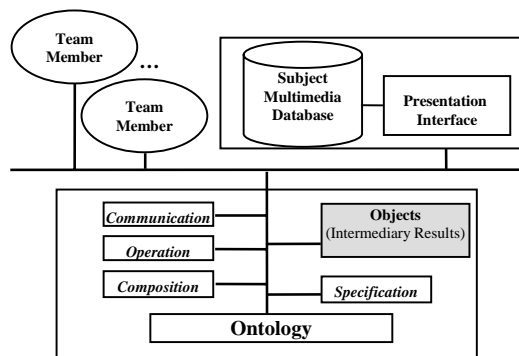


Figure. 5 System Structure of Heritage Archiving Mediator

HAM 提供進行數位典藏工作中的使用者彼此溝通的管道(communication)，此外使用者可經由平台加以操作與典藏數位物件(operation)，同時平台對於處理數位物件設計了完整的規格(specification)，另外描述了數位物件的傳遞與整合(composition)。

HAM 的運作是基於團隊合作，並且透過角色扮演得知自身於工作流程所需實作的任務，了解目前的工作進度，以及前後任務的關聯性。以 Figure. 6 為例，每個工作人員都能看到即時的輸出物件，並且知道 subject domain specialist 是該 activity 正在對於該物件進行審核，決定是否 approve 進入下一個工作階段。同時成員之間還可以利用討論區，加強彼此間的互動，共同解決典藏工作進行中的問題。(Figure. 7)。另外每個工作都有其需要負責的 worker，HAM 提供不同性質的工作的畫面讓 worker 使用。以 Figure. 8 為例，data management specialist 正針對之前傳遞進來的物件，進行資料加註的動作。



Figure. 6 A View Panel to Evaluate Results

5. 實例應用

HAM 在政治大學一共有三次專案的實際使用。第一次是由資訊科學系與廣電系合作，開發顧正秋藝術網站。本專案透過 HAM 並且結合五種專家的合作，用了兩個月的時間，完成



Figure. 7 A Discussion Panel to Exchange Comments



Figure. 8 A Work Panel to Produce Output

資料庫的建立與使用者介面的呈現 (<http://koo.theatre.nccu.edu.tw>)，資料庫蘊藏的資訊相當龐大，有 61 筆事件年表，84 份作品，294 篇相關報導，325 章圖片以及 40 分鐘以上的影片與 15 分鐘以上的聲音。(Figure. 9)

第二個專案是政治大學傳播學院影像資料庫 (http://godel.cs.nccu.edu.tw/news_photo/cocid/index.asp)，目的是將政治大學創校以來的老照片數位化、加註化與文件化，主要參與專案的人員以學生為主，然而本次的專案時間用了三個月就實作出多媒體照片資料庫，其中包含 3000 張以上的照片以及 35 篇的主題報導，這個資料庫已經被視為政治大學的重要資產。(Figure. 10)

第三個專案是政治大學圖書館與檔案人物網站 (<http://godel.cs.nccu.edu.tw/dl/index.html>)，專



Figure. 9 顧正秋藝術網站



Figure. 11 政治大學圖書館與檔案人物



Figure. 10 政治大學傳播學院靜態影像資料庫

案目標是實作圖書館界與檔案界的人物典藏，包含訪談紀錄、作品、報導、記事年表、檔案照片與相關連結等等，總共歷時一個學期的時間。這次的專案典藏總共做了 21 個人物，透過自行設計的 metadata，將所有資訊都納入資料庫之中。(Figure. 11)

6. 結論

多媒體的主題式建檔與內容建置對於我們的文物遺產保存有不可磨滅的地位，其中能夠掌握文化尚存的時間性最大的關鍵就在於廣泛的納入業餘人士的力量，讓各人發揮自己的專業技能，投入到數位典藏的領域之中。我們以資源導入的觀點，發展出數位典藏建置分工合作模

式與工作流程架構，實做了以 ontology 為驅動力量的 HAM，該系統提供了整合業餘資源之平台，並輔助數位典藏工作之進行；同時 HAM 還經過三個專案的測試，藉由實際資料的印證，增進系統的功能。我們相信，以 HAM 為基礎的數位典藏工作將提供給相關工作者新的思維與成就。

7. 參考文獻

- [1] M. R. Lyu, E. Yau and S. Sze, "A Multilingual, Multimodal, Digital Video Library System", *Proceedings of the second ACM/IEEE-CS joint conference on Digital libraries*, pp.145-153, 2002.
- [2] I. Witten, M. Loots, M. Trujillo and D. Bainvridge, "The promise of digital libraries in developing countries", *Communications of the ACM*, 44(5), pp.82-85, 2001.
- [3] B.N. Schilit, M.N. Price and G. Golovchinsky, "Digital Library Information Appliances", *Proceedings of the third ACM conference on Digital Libraries*, pp.217-226, 1998.
- [4] C. Lagoze, D. Fielding and S. Payette, "Making Global Digital Libraries Work: Collection Services, Connectivity Regions, and Collection Views", *Proceedings of the third ACM conference on Digital Libraries*, pp.134-143, 1998.
- [5] R. Rao, J. O. Pedersen, M. A. Hearst, J. D. Mackinlay, S. K. Card, L. Masinter, P. Halvorsen and G. G. Robertson, "Rich Interaction in the Digital Library", *Communications of the ACM*, 38(4), pp.29-39, 1995.
- [6] D. Charitos, G. Lepouras, C. Vassilakis, V. Katifori, A. Charissi, L. Halatsi, "Designing a virtual museum within a museum", *Proceedings of the 2001 conference on Virtual reality, archeology, and culture heritage ACM Press*, pp.284-284, 2001.