

科技部補助專題研究計畫成果報告 期末報告

數位典藏長期保存策略及系統建置之研究：以中研院台史
所檔案館為例

計畫類別：個別型計畫
計畫編號：MOST 102-2410-H-004-220-
執行期間：102年08月01日至103年09月30日
執行單位：國立政治大學圖書資訊與檔案學研究所

計畫主持人：薛理桂

計畫參與人員：碩士班研究生-兼任助理人員：黃若瑜
碩士班研究生-兼任助理人員：林育如
博士班研究生-兼任助理人員：陳慧嫻

報告附件：出席國際會議研究心得報告及發表論文

處理方式：

1. 公開資訊：本計畫涉及專利或其他智慧財產權，1年後可公開查詢
2. 「本研究」是否已有嚴重損及公共利益之發現：否
3. 「本報告」是否建議提供政府單位施政參考：否

中華民國 103 年 11 月 20 日

中文摘要：

中文關鍵詞：

英文摘要：

英文關鍵詞：

數位典藏長期保存策略及系統建置之研究：以中研院台史所檔案館為例

壹、緒論

數位典藏與數位學習國家型科技計畫至今已執行十年，產生了豐富的數位化成果，在電腦科技與儲存媒體迅速進步的情形下，如何將數位化成果透過模擬、轉置等方法長期保存豐碩的成果已成為重要議題。本研究擬藉由文獻探討，瞭解國際上數位保存策略、相關系統及工具，首先以曾參與數位典藏國家型科技計畫的重要典藏單為焦點訪談對象，以了解進行長期保存時所遭遇的問題和困難。再以 Archivematica 作為實驗工具，以中央研究院台灣史研究所檔案館為測試實例，期能提出一個經濟實惠的數位保存策略及作業模式，作為國內檔案相關單位參考之依據。

貳、研究計畫之目的與研究問題

一、研究目的

本研究希望針對現有數位檔案保存的問題，提出改善方案，與數個重要典藏單位進行焦點座談，後以中央研究院台灣歷史研究所(以下簡稱台史所)檔案館之數位檔案保存性後設資料系統為個案，以系統實作方式，並評估成效，以達到提昇館藏品質及行政效能之目標，進而提供其他檔案館進行數位保存後設資料系統設計時之參考。本研究希望達到以下三大目的：

- (一) 探討國內典藏單位數位檔案保存之政策與作業方式。
- (二) 探討導入 **Archivematica** 軟體對中研院台史所檔案館保存政策及人力作業之影響。
- (三) 探討數位保存標準模式/流程如何建立與執行。

二、研究問題

根據上述研究目的，本研究主要探討的問題有下列幾項：

- (一) 探討數位保存後設資料標準之內涵、特性及世界各國應用之現況以及所遭遇的困難與限制為何？
- (二) 探討中研院台史所檔案館在管理數位檔案之政策與人力執掌、工作執行現況為何？
- (三) 導入 PREMIS 標準及系統工具後，對中研院台史所檔案館的保存政策及人力之影響為何？
- (四) 理想的數位保存標準模式的建置與流程為何

參、文獻探討

一、長期保存後設資料標準 PREMIS

(一)背景資訊

PREMIS 最初由 OCLC (Online Computer Library Center, OCLC) 與 RLG (Research Libraries Group, RLG) 贊助的 Preservation Metadata Framework 工作小組所發展，在 2001-2002 年該小組提出報告—A Metadata Framework to Support the Preservation of Digital Objects (the Framework)。2003 年 6 月，在 OCLC 與 RLG 的共同贊助下，集合全球圖書館、博物館、檔案館、政府，及私人單位的數位保存專家，成立了 PREMIS (PREservation Metadata: Implementation Strategies, 以下簡稱 PREMIS) 工作小組。工作小組的任務是發展數位保存的核心後設資料，以協助數位資料的創造、管理及利用。(PREMIS Editorial Committee, 2012)

PREMIS 工作小組的目標是 (OCLC/RLG, 2005, p.vii)

- 定義一套可執行的「核心」保存後設資料元素，能廣泛應用於數位保存社群。
- 設計資料字典以支援核心保存後設資料元素集。
- 檢驗或評估在數位保存系統中，保存性後設資料的編碼、儲存與管理以及在系統間交換保存性後設資料等策略。
- 執行試驗計劃在多樣化的系統設定中測試該小組的建議與最佳實踐。
- 探索合作生產與分享保存性後設資料的機會。

為了找出保存機構如何實際執行保存性後設資料，於 2003 年 11 月調查 70 個與數位保存相關的組織。這個調查提供了解數位保存目前發展水平的機會，草擬問題引出關於政策、管理與資金、系統架構與保存策略，以及後設資料實施的資訊。PREMIS 工作小組在 2005 年 5 月完成此任務，並發表「保存性後設資料資料字典：PREMIS 工作小組最終報告 (Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group)」，此為 PREMIS v1.0。(PREMIS Editorial Committee, 2012)

此資料字典定義保存性後設資料為：(PREMIS Editorial Committee, 2012)

- 支援可行性、執行性、可理解性、可信賴性以及在保存性數位物件內容的一致性。
- 為了盡可能表現資訊，保存典藏單位必須知道如何長期保存數位素材。
- 強調保存性後設資料是：嚴厲的、清晰的，遵守指導方針來建立、管理與使用，及指向自動工作流程。
- 具體化科技中立，沒有對於保存性科技、策略、後設資料儲存及管理有任何假定 (assumptions)。

此外，PREMIS 工作小組也出版了一組 XML schema 來支援數位文件系統中的資料字典。PREMIS 資料字典贏得了 2005 年英國保存大獎 (British Conservation Awards) 之數位保存獎，同時也贏得了 2006 年美國文件檔案人員學會之保存出版品獎(American Archivists Preservation Publication Award)。(PREMIS Editorial Committee, 2012)

在 2005 年發行資料字典後，PREMIS 工作小組結束其任務，改由國會圖書館贊助運作 PREMIS 維護活動，並開始其他工作以進一步瞭解保存性後設資料及相關主題。除提供資料字典、XML schema 和相關資源的永久的網站外，維護活動同時也運作 PIG (PREMIS Implementers Group) 討論清單及協作平台，提供資料字典的使用指導，並授權關於保存性後設資料的研究主題。維護活動並建立編輯委員會，負責未來資料字典及 XML schema 的發展，同時推廣它們的使用。編輯委員會成員背景包含多個國家與機構。(PREMIS Editorial Committee, 2012)

在資料字典發行的同時，也決定暫停其內容修改至少 18 個月，讓數位保存委員會閱讀並且消化、試驗其操作性、除錯，以及最重要的是提供關於資料字典改進、增加價值及簡單運用的回饋。在此期間接收到許多機構的相關回饋，在 2007 年時，編輯委員會累積到一定程度的評論，因此決定進行第一次的資料字典版本更新。編輯委員會的成員進行資料字典的修訂，在重新修訂的過程中致力於滿足所有利益相關者。委員會討論保存社群的相關議題，並徵集對於改版的建議，及諮詢外部專家的意見，PREMIS Data Dictionary for Preservation Metadata 第二版就在此過程中誕生。(PREMIS Editorial Committee, 2012, p.2)

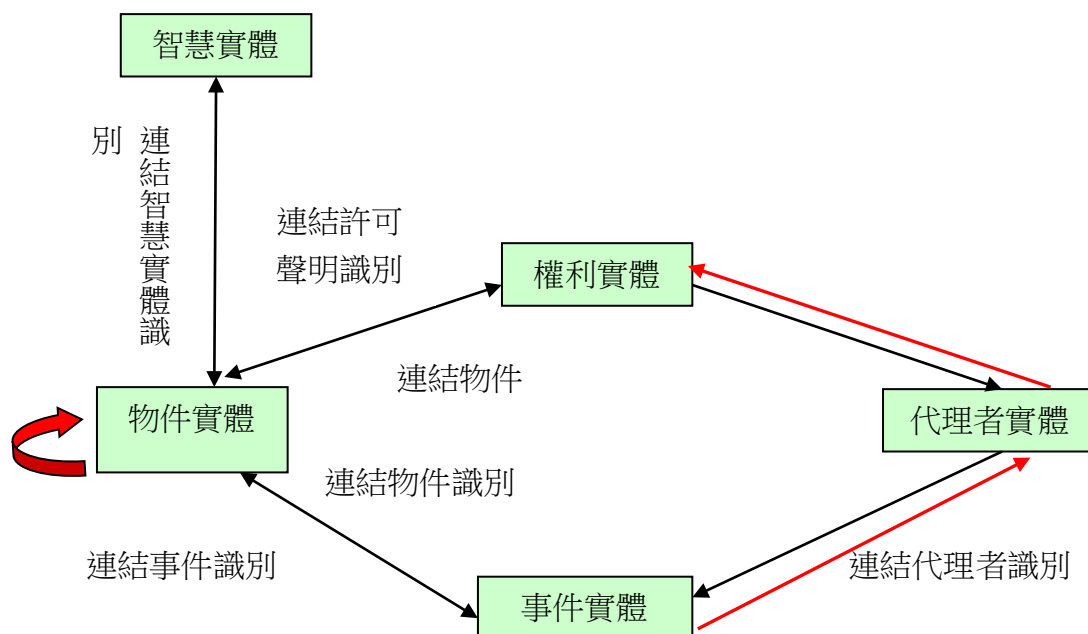
本標準的 Framework 主要以開放式檔案資訊系統(OAIS)參考模組(ISO14721)所建置。OAIS 資訊模組提供一概念性的基礎，提供資訊物件與典藏物件的封包和他們相關的後設資料的架構的分類。Framework 可視為 OAIS 資訊模組的精心之作，透過保存後設資料對應其概念性架構來闡述。PREMIS 可視為將 Framework 轉化成資料字典中一組可執行的語意單元。然而，PREMIS 語意單元比 OAIS 定義提供的要更具體，其期望從概念性架構進展至實際可運作的架構。(PREMIS Editorial Committee, 2012)

(二) PREMIS Data Model

為促進 PREMIS 後設資料元素的邏輯組織，本標準發展了一簡單的模組，其包含在數位保存活動中有關的五種類型的實體：智慧實體(intellectual entities)、物件實體(object)、事件實體(event)、權利實體(rights)、代理者實體(agents)。在資料模型圖中，實體都以方型表示，他們彼此間的關係則以線表示。箭頭表現

的方向即是資料字典中所定義的關係，例如：箭頭自權利實體至代理者實體意味著權利實體的後設資料包含語意單元，以辨別相關的代理者實體。雙箭頭意味著相互連結的關係，見圖 1，並敘述如下：(PREMIS Editorial Committee, 2012)

圖 1 PREMIS 資料模組關係圖



資料來源：(PREMIS Editorial Committee, 2012)

1.智慧實體(intellectual entity)

一組連貫的內容其可被描述為一單元，例如，一部特別的書、地圖、照片或是資料庫。一智慧實體可包含其他智慧實體，例如一網站可包含一網頁，一網頁可包含一張照片。一智慧實體可能有一個或一個以上的數位表示法。(PREMIS Editorial Committee, 2012)

2.物件（或數位物件）實體(object or Digital Object)

數位形式中資訊分離單元。物件實體細分為表示法、檔案、位元流，說明如下：(PREMIS Editorial Committee, 2012)

- 表示法：代表或具體化內容(intellectual)實體的數位物件。表示法是一組已儲存的數位檔案及結構化後設資料，提供內容(intellectual)實體一完整且合理的詮釋(rendition)
- 檔案：為作業系統所知已命名且有次序的位元組。
- 位元流：在檔案中的連續或非連續性資料，其具有保存目的屬性資訊。

3.事件實體(event)

一動作包含或影響至少一個物件或保存機構已知的代理者 (PREMIS Editorial Committee, 2012)

4.代理者實體(agent)

與一物件的生命週期的事件中相關的人、組織或軟體程式系統或與物件相關的權利 (PREMIS Editorial Committee, 2012)

5.權利實體(right)

關於一物件或代理者的所有權或許可權的聲明 (PREMIS Editorial Committee, 2012)

(三)PREMIS 物件關係

典藏庫中的物件可與典藏庫中一個或更多物件有相關。PREMIS 利用語意單元來支援物件間關係的敘述。PREMIS 工作組利用收集現有保存性後設資料計畫的範例來探索,發現廣泛的關係表示,例如「從...移轉(is migrated from)」、「是...縮圖(is thumbnail of)」等等。此外,工作組也檢視都柏林核心集(Dublin Core),關聯(Relation)欄位,如是..一部份(IsPartOf)、是...的格式(is format of)等,總結在物件關係中大多是三種基本關係的變異:結構的、來源的、附屬的,詳述如下:(PREMIS Editorial Committee, 2012)

1.結構關係(Structural relationship)

結構的關係表現在物件間的部分關係。在檔案間的關係構成智慧實體的表現是必要的保存後設資料。如果一保存典藏庫不能將數位物件回復原狀,則其未保存該物件。例如:單一的數位物件(例如一張照片)結構資訊是很小的,由檔案構成表示法。其他數位物件例如電子書和網站可能會有相當複雜的結構關係。

2.來源關係(Derivation relationship)

來源關係產生於一物件的複製或轉製。所產生的物件的智慧實體是相同的,但是物件的形式等可能是不同的。當 X 形式的 A 檔案轉製產生 Y 形式的檔案 B,則檔案 A 與檔案 B 間有來源關係。

許多數位物件是很複雜的,由於保存活動的關係使結構與來源資訊歷經時間改變。例如,一本數位化的圖書可能由 400 頁 TIEF 影像檔呈現,之後可能轉製成四個 PDF 檔案,每一檔案包含 100 頁。

物件間的結構關係在物件被典藏庫收入前能由來源動作所建立。例如,一個文字處理的文件能分別被製成 PDF 檔案與 XML 檔案。如果僅 PDF 和 XML 檔案

提交至保存機構，這些物件是相同智慧實體的不同表現，與來源文字處理檔案有母子關係。他們彼此間並沒有來源關係，但有結構關係如同血緣關係。(p.13)

3.附屬關係(A dependency relationship)

附屬關係存在於當一物件需要其他者來支援其功能、傳遞或內容的連貫性。一物件可能需要字型、樣式表、DTD、圖解或其他檔案，其不是該物件本身正式的一部份，但卻是必要的。資料字典將附屬關係視為環境資訊的一部份，在語意單元中為 **dependency and swDependency**

(四)不同類型實體間的關係

資料模組圖示(如圖 1)以箭頭顯示不同類型實體間的關係。物件與相關的智慧實體，物件與相關的事件，代理者與相關的事件等等。PREMIS 表現關係如連結資訊般，如實體 A 的資訊指向相關的實體 B。在資料模組中的每一個實體都有獨特的識別碼作為指向。例如，物件實體有箭頭指向智慧實體和事件實體。在 PREMIS 中，這些都由語意單元 **linkingIntellectualEntityIdentifier** 和 **linkingEventIdentifier** 表示。(PREMIS Editorial Committee, 2012, p.14)

PREMIS 實體間關係：一比一原則(The 1 : 1 principle)詳述如下，(PREMIS Editorial Committee, 2012)

在數位保存中，以儲存的物件產生新的副本或版本是很普遍的情況。例如，前面轉製 X 形式的檔案 A 在一程式中輸入後可能輸出成 Y 形式的檔案 B。檔案 A 與 B 有兩個考量，考量之一為他們兩個都是一單一的物件，其歷史包含從 X 轉製成 Y，或是將他們看成是兩個不同的物件，其中的關係由轉置事件所產生。

1 : 1 原則在後設資料顯示每個描述僅描述一種資源。如同應用 PREMIS 後設資料，在保存典藏庫中擁有的每一物件實體(檔案、位元流、表示法)被描述為一組靜態的位元。改變一個檔案(或位元流或表示法)是不可能的，僅能產生一個新的檔案(或位元流或表示法)其與來源物件相關。例如，檔案 A 與 B 是不同的物件，他們之間有來源關係。資料字典對於物件的產生日期有一語意單元 (**dateCreatedByApplication**)，但對於一物件的修正日期則沒有，因為一物件在定義上是不能被修正的。

當新物件來自於現存的物件，產生新物件這個事件應被紀錄為一事件，其將會有一日期/時間標記。物件間的關係應被紀錄，使用與物件實體相關的 **relationship** 語意單元。

(五) PREMIS 的限制

1. 資料字典的重心不在描述性後設資料

描述性後設資料是用來描述智慧實體，例如智慧實體的產生者、產生日期、描述、格式、權利、主題與關鍵字等，而保存性後設資料主要描述保存智慧實體的軟硬體相關資訊、保存日期、保存方法等。所有保存典藏庫都會包含描述性後設資料或是連結在典藏庫以外的描述性後設資料。描述性後設資料非重心的原因有二：第一，描述性後設資料已有多個標準。MARC、MODS、都柏林核心後設資料集、EAD 等都是描述性後設資料標準。第二，描述性後設資料通常是特殊的領域。在國際上，目前多使用 METS 標準結合智慧實體的描述性後設資料與保存性後設資料。(PREMIS Editorial Committee, 2012)

2. 代理者(Agents)

PREMIS 並沒有定義代理者的特性。後設資料描述人們、組織和其他實體能作為代理者，已在許多現存的格式和標準中定義，例如 MARC、vCard、MADS 和其他正在發展的綱要。只要保存機構能正確地辨別代理者，其他的代理者特性將由需求決定。(PREMIS Editorial Committee, 2012)

3. 權利(Rights)

PREMIS 僅定義權利的特性和保存活動相關的許可，與取用或散佈無關。修訂版擴增權利資訊的語意單元，並允許延伸使用外在權利後設資料綱要。(PREMIS Editorial Committee, 2012)

4. 技術性後設資料(Technical metadata)

技術性後設資料描述數位物件的實體而非智能特性。詳細的、格式特定的技術性後設資料在執行大多數的保存策略時，是必要的，但是工作組既沒有時間也沒有專家，可獲取不同類型的數位檔案特定格式技術性後設資料。因此，在資料字典中有限的技術性後設資料，其相信可應用於任何格式的物件。技術性後設資料的更進一步發展將留給格式專家。延伸性機制為提供語意單元 `objectCharacteristicsExtension`，其可用於外在的技術性後設資料綱要。(PREMIS Editorial Committee, 2012)

5. 媒體或硬體細節

工作小組並未試圖定義媒體或硬體詳細的使用說明的後設資料。例如，PREMIS 為辨別一物件儲存的媒體，定義一語意單元。保存機構將可能想知道更多有關使用媒體的詳細資訊。例如，典藏庫將資料儲存在 DVD 上，可能需要知道特定 DVD 單元的特殊技術特性，例如製造廠商、染色素材和染色厚度。PREMIS 將描述媒體與硬體特性的後設資料的定義留予該領域的專家。(PREMIS Editorial Committee, 2012)

6. 業務規則(Business rules)

工作小組並未試圖描述典藏庫的業務規則，雖然此種後設資料在典藏庫中的保存來說是必要的。業務規則編纂保存策略的應用和紀錄典藏庫的政策、服務、任務和角色。保存期限、處置、危機評估、持久率、媒體更新時間表等等都是與物件相關的但並非是物件的實體特性。唯一例外是保存層級可紀錄於物件 (preservationLevel)，因為這對任何保存典藏庫都是關鍵的資訊。更多的業務規則可透過定義規則實體(Rules entity)，其近似權利實體，但目前並未包含於現在這個版本內。(PREMIS Editorial Committee，2012)

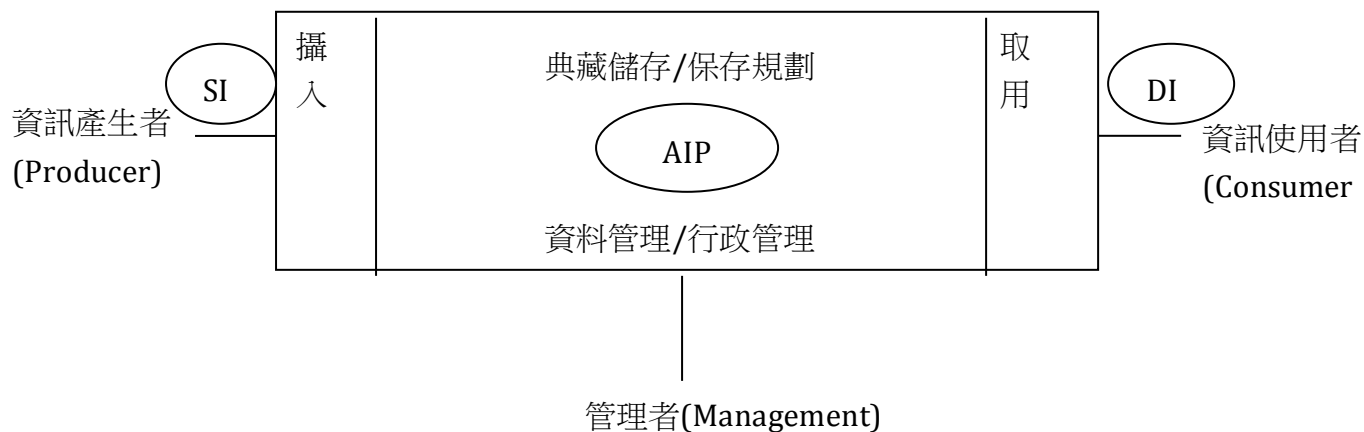
二、OAIS 參考模組

開放性檔案資訊系統參考模式(Refernce Model for Open Archival Information System，簡稱 OAIS)是國際標準組織(International Standards Organization，簡稱 ISO)所屬 NASA 的太空資料系統諮詢委員會(Consultative Committee for Space Data Systems，簡稱 CCSAS)於 1999 年所發展的一套標準模式，並於 2001 年提出修正版。OAIS 模式是一個典藏系統的概念模型，可應用於任何型態資料，尤其是數位資料的長期保存並提供檢索取用。它描述典藏機構所處的環境、典藏機構本參的功能組成、支援典藏機構工作流程的資訊基本架構等，同時也為對於數位典藏有興趣的不同領域人士，提高一種高層次的討論語言。應用 OAIS 的機構，可以得到國際標準的種種好處，藉著使用共同的參考模式、共同用詞及共同的概念結構，大家更容易分享想法及交換經驗。(王文英，2003)

OAIS 是「一個人與系統結合而成的組織，負責為特定社群保存資訊並提供使用」，它同時定義數位典藏機構的功能模型及資訊模型。在 OAIS 描述的功能模型中，數位典藏機構應具備的功能包括：檢索取用、行政管理、資料儲存、資料管理、資料獲取及保存工作規劃，這些功能與資料產生者、使用者及館方的互動，形成 OAIS 的功能模型。(王文英，2003)

依據 OAIS 參考模式的定義，可將數位典藏活動視為一個由 OAIS 典藏系統本身，加上資訊產生者(producer)、資訊消費者(consumer)及管理部門(management)等四個實體所組成的互動環境，如圖二所示。而 OAIS 典藏系統又包含攝入(Ingest)、典藏儲存(archival storage)、資料管理(data management)、行政管理(administration)、保存規劃(preservation planning)、取用(access)等六種主要功能實體。此外，OAIS 參考模式中數位資訊係以資訊封包(Information package)的形式存在；而資訊封包依任務不同可分為三種：1.由資訊產生者傳送給 OAIS 典藏系統的「投入資訊封包」(Submission Information Package，簡稱 SIP)2.儲存於 OAIS 典藏系統中的「典藏資訊封包」(Archival Inforamtion Package，簡稱 AIP)3.當資訊使用者發出取用要求時，從 OAIS 典藏系統回應給

使用者的「遞出資訊封包」(Dissemination Information Package, 簡稱 DIP)。(林信成、黃文彥, 2012)



三、長期保存政策發展現況與要項分析

Sheldon(2013)於 2013 年出版數位保存政策規劃的報告，其研究發現針對 2008-2013 年之間 33 所機構中，檔案館佔 48%，11 所為政府機構、5 所為學術機構。圖書館佔 45%，其中 8 所為學術機構，7 所為政府單位。而博物館僅佔 2%，1 所為政府機構，1 所為非營利組織。以出版年份而言，2008-2010 年之間出版 14 份文件，2011-2013 年間則有 19 份文件，從 2011 年後呈現急遽成長趨勢。(Sheldon, 2013)

Sheldon 重新分析 33 份政策後，提出 19 類標準，包含取用與使用、登錄/攝入、稽查(audit)、參考書目、合作(Collaboration)、內容範圍、詞彙/術語、命令(Mandates)、後設資料/文件(Documentation)、政策/策略評估探討、保存模組/策略、保存規劃、權利和限制管理、角色與責任、安全管理、選擇/鑑定、員工訓練/教育、儲存、複製和備份、永續性規劃等。最常使用的標準包含保存策略/模組、合作以及內容範圍。最少被提到項目為：登錄/攝入、稽查和保存規劃。(Sheldon, 2013)

數位保存的管理方面，Sheldon (2013) 提到國會圖書館的機構發展主管認為，電子文件管理及先前處理數位資料的經驗，直接影響檔案館、圖書館和博物館的數位管理方式。1960 年代圖書館建置 MARC 機讀格式，使圖書館可在虛擬環境中維護其書目資訊內容，並促使檔案館和博物館在幾年後也追尋相似的程序。當檔案館和圖書館開始在機構中收集越來越多的數位內容時，管理電子文件的經驗使他們有能力保存數位資料。

就數位保存經驗方面，博物館著重收集某些時間為主的媒體，維護由原件和數位資料組成之混合式物件，然而檔案館和圖書館則是以數位化或原生數位館藏為主。紐約 Rhizome 博物館數位保存人員 Ben Fino-Radin 曾說，在規劃數位保存政策時，檔案館和圖書館是領先於博物館，乃因多數博物館多投入媒體藝術作品的數位維護(conservation)。博物館的工作和圖書館、檔案館略有不同，以維護媒體為主之藝術作品的實體和內容，以及藝術家的原意，需要特定維護措施。(Sheldon, 2013)。因此，圖書館、博物館和檔案館這三類機構，在數位長期保存管理或策略上，有互相學習、相輔相成的關係。

在保存政策所包含項目中，檔案館方面較特別的項目為詞彙/術語、安全管理、儲存、複製和備份、政策/策略評論以及權利和限制管理。圖書館方面則是較著重合作、參考書目、永續規劃以及後設資料/文件。博物館方面的政策則較為簡略。(Sheldon, 2013)

(一) 長期保存政策之目的

英國國家檔案館出版「數位保存政策：檔案館指南」中提到，數位保存政策目的意在表明為何典藏單位要進行數位保存，如下所示：(The National Archives, 2011)

- 保存數位文件的目的與動機。
- 數位保存如何能使檔案館達成其核心目的（例如作為一個記憶機構）。
- 數位保存如何與其母機構的策略目標保持一致（例如檔案館和地方政府）。
- 提高長期取用資料的效益，透過數位保存滿足經濟、名譽、再使用和文化危機方面的需求

數位保存政策對於執行數位保存時也扮演權威性的角色，擔任連接典藏單位與外部團體（例如合約者、文件的寄存者或捐贈者）的角色與責任，如下所述：(The National Archives, 2011)

- 定義一組簡潔的成功標準，使之可作為對照。
- 定義數位保存活動的範圍，包含廣泛的文件種類。
- 定義檔案館何時以及如何鑑定數位文件，這些數位文件必須具有歷史價值且符合技術特性才得以典藏。
- 辨識數位保存策略的存在（以及擁有者）。
- 指出在此數位保存策略中的任何標準。

以上項目都是檔案館以一致的框架來實現數位保存，不需再為其他資源或服務來修正需求，因在政策中都已明述。政策也提供清楚的責任界線以便典藏單位解決爭議。本文以兩所國家檔案館與一所地方檔案館所訂定之數位保存政策目的（見表 1），作為檔案館訂定數位保存政策之參考。

表 1: 長期保存政策之「目的」範例

原始出處	長期保存政策之目的
英國國家檔案館保存政策 (The National Archives, 2009,p.3)	長期保存政策的目的是陳述並引導英國國家檔案館館藏之保管與保存的原則，提供英國國家檔案館在決策時全面性框架，並作為程序發展的參考，本政策是關於數位文件及複製品的所有格式與材料。
澳洲國家檔案館數位保存政策 (National Archives of Australia, 2011)	<p>政策聲明：澳洲國家檔案館確保澳洲聯邦檔案資源的長期保存和可用性，這些檔案資源是以數位格式產生或管理的。</p> <p>政策目標如下：</p> <p>澳洲國家檔案館的目標是保存任何類型的數位文件：</p> <ul style="list-style-type: none"> ● 以任何類型的應用軟體產生的數位文件 ● 在任何電腦平台所產生的數位文件 ● 以任何數位媒體所傳遞的數位文件 ● 從任何澳洲聯邦機構或眾所公認的個人文件寄存者（個人文件是現任和歷任總理、部長以及其他與聯邦政府相關的人的文件） ● 提供現在和未來查詢與取用
赤郡檔案館及地方研究的數位保存政策 (Cheshire Archives & Local studies, 2010)	數位保存政策的目的是規劃如何達成保存數位材料的方式，並釐清寄存者和其他典藏管理者預期從數位保存的服務中獲得什麼。這個政策也解釋以目前的資源、基礎建設和可取得的技能中，哪些服務是無法執行或保證的。

(二) 數位保存政策與數位保存策略關係

英國國家檔案館於 2011 年出版長期保存政策指南，提供檔案館發展長期保存政策參考。指南中提到政策和策略須確保可驗證及可信賴的方法，以保存數位文件的完整性。依據數位保存的規模和複雜度，以小型組織為例，若須政策中部份的策略，只要確認策略的改變，而不須要修改政策。(The National Archives, 2011)

數位保存策略廣泛由典藏單位所採用的方法，確保數位文件的內容保留於歷時可用的形式。策略工作由辨識特定的時間與工具，將會決定數位保存活動的發生。例如：(The National Archives, 2011)

- 移轉：檔案格式接近保存期限時，將移轉至更長久的格式。
- 攝入：當文件的檔案格式歷經定義好的正規化程序後，以更長久持續的格式呈現文件的內容。

然而，數位保存政策必須能辨識保存策略和其擁有者。若強調策略和所定義的保存活動，僅須要允許政策保持獨立性，不須要更新所造成的改變。表 2 顯示數位保存之政策與策略關注的不同焦點。此為英國國家檔案館所舉之範例，非政策或策略完整應包含的內容。(The National Archives, 2011)

表 2:英國國家檔案館政策與策略內容比較表

功能	政策	策略
數位保存的擁有權	●	
數位保存的策略聯盟	●	
數位保存的範圍	●	
角色和責任	●	
與其他政策的關係	●	
與其他文件的關係	●	
數位文件儲存的類型		●
接受/優先的格式		●
移轉的時機		●
移轉的類型		●
正規化的時機		●
正規化的類型		●

(三) 數位保存政策的範圍

在數位保存政策範圍方面，英國國家檔案館認為檔案館將從不同的來源得到數位文件。因此，政策必須辨識其範圍，集中於數位文件相關的徵集和館藏政策。必須要包含下列的陳述：(The National Archives,2011)

- 作為保存之數位文件的類型（包含內容、結構和格式）。
- 從哪些機構或組織蒐集的文件。
- 這些文件的位置（指儲存和系統方面）。

數位保存政策清楚地詳述保存數位文件之蒐集和辨識的程序，則可優先確保數位文件的館藏和維護是在政策的規範之內。同時檔案館也可拒絕典藏定義範圍外的文件，例如一批館藏難以保存或超出保存的能力時。其他的可能作法是僅保存館藏的特性，或建議他們寄存至其他檔案館。

在此，本文以英國國會檔案館保存政策與澳洲國家檔案館數位保存政策所訂定之範圍（見表3），作為檔案館制定數位保存政策之參考。

表 3: 長期保存政策之「範圍」範例

原始出處	長期保存政策之範圍
<p>英國國會檔案館保存政策 (Parliamentary Archives, 2009)</p>	<p>本政策應用至國會檔案館保管的數位形式資訊資源，可細分為三大類</p> <ol style="list-style-type: none"> 1. 「原生數位」資源，因業務目的以電子形式產生與管理 2. 「人造數位」資源，以非電子的形式產生，但之後因下列目的而轉置為數位形式： <ul style="list-style-type: none"> ● 業務，在功能上等同於第一類 ● 保存，因原始保存形式無法使用，例如由於實體損壞或技術過時。這些在功能上等同於第一類。 ● 取用 3. 「再生」數位資源，以數位形式產生，因業務目的以非電子形式管理（例如，在「印出為紙本」的政策下），但之後因業務、保存或取用目的再數位化。這些在功能上與第二類的相同。 <p>數位資源包含數位文件和其他數位素材。數位資源可被視為混合實體，包含內容和描述部分。</p> <ul style="list-style-type: none"> ● 內容由原始應保存的物件構成 ● 描述的部分由後設資料描述構成或其他與內容相關的部分。 <p>這些部分都是資訊物件，可能以位元流、位元流的一部分或在電腦檔案系統中的一組位元流呈現。實體上，後設資料可能與內容整合，或分開。本政策因此提供保存資訊內容物件和他們相關的後設資料，維護兩者間的一致性，並產生新的後設資料記錄所執行的保存程序</p>
<p>澳洲國家檔案館數位保存政策 (National Archives of Australia, 2011)</p>	<p>依據 1983 澳洲檔案法，澳洲國家檔案館負責保存聯邦政府的文件，以此形成聯邦政府的檔案資源。這其中包含原生數位文件和傳統原件的數位保存典藏副本（master copies）。</p> <p>國家檔案館僅接受來自政府機構的文件，並且這些文件經由文件權威（Records authorities）判定保留作為國家檔案（National Archives）（也就是永久保存），（「判定」是使用文件權威或其他工具決定哪些要保留、銷毀或移轉文件的一個</p>

原始出處	長期保存政策之範圍
	<p>程序。文件權威是由國家檔案館和機構特定核心業務公佈給個別機構的。)</p> <p>在例外的情況下，澳洲國家檔案館將接受未經辨識文件價值的文件，例如具有風險的文件或被認為是重要的資源。</p> <p>澳洲檔案館不保存用來產生、管理或呈現數位文件的方法，例如文件管理軟體。澳洲國家檔案館接受從系統中輸出的數位文件和他們的後設資料，但不輸出系統本身。</p>

(四) 數位保存政策的責任與角色

英國國家檔案館數位保存政策指南中提到定義角色和責任是必要的，以此辨識典藏單位中誰應負責管理和保存數位文件。

這個政策辨識誰應該負責數位保存，以及對他們的期待。如果是非真實的期待，將會使保存真實數位文件的目的落入風險。為避免如此，在政策發展的期間，應諮詢所有可能涉入的人。這將使典藏單位評估自己或預期的合作夥伴是否真實可支援數位保存。指南將責任和角色分為三大類：(The National Archives, 2011)

(1) 數位保存的擁有權

若要使保存政策發揮效能，保存政策在典藏單位中必須擁有高階地位，使數位保存政策可進行策略決定，並對數位保存有長遠影響。

(2) 服務提供者與合約者

典藏單位可能需要外來的協助，以執行數位保存活動，彼此間的關係需要清楚辨識，並期待服務層級具正式同意和記錄，以確保所有合作夥伴的可靠性。當採用服務提供者時，很重要的是委託之典藏單位仍保留負責數位文件。

(3) 寄存者/捐贈者

數位文件的寄存者/捐贈者將需要確保文件以適當的狀況提交至檔案館，政策需要陳述接受寄存之數位文件時的需求，包含提供數位保存相關文件（例如可接受或優先的檔案格式清單，以及最小描述和技術性後設資料）等相關的文件，以及廣泛的徵集或館藏問題（例如文件是否符合檔案館徵集和館藏政策的標準）。

本文以英國國會檔案館保存政策所訂定之責任和角色（見表 4），作為檔案館制定數位保存政策之參考。

表 4: 長期保存政策之「責任與角色」範例

原始出處	長期保存政策之責任和角色
英國國會檔案館保存政策 (Parliamentary Archives, 2009)	執行數位保存時，隨著時間變革需要與國會外部的單位合作。廣泛而言，國會檔案館將負責設立、維護及監控、數位保存策略和政策、徵集及館藏政策、技術監控(Technology Watch ,with PICT)、保存規劃、保存行動(with PICT)、及呈現(with HC 及 HT 資訊服務)。PICT 將確保 ICT 策略及所有的 ICT 計畫都依循本政策及相關的標準。內容產生者將負責確保數位資源的產生與管理都依據本政策及相關的標準。詳細的角色與責任將是數位保存計畫的一部份。

(五) 數位長期保存政策和其他政策的關係

數位保存政策在執行和實現數位保存活動時，必須考量的其他政策。詳述這些政策之間的關係，並指出其他政策的優先順序。英國國家檔案館數位保存政策指南，以表列出組織可能已建立的重要政策(詳表 5)，對於數位保存將有影響，因此必須作為數位保存政策的參考。(The National Archives, 2011)

表 5: 長期保存政策與其他政策之關係表

政策	關係
接受/優先的檔案格式的清單	檔案館將接受的軟體檔案格式清單(例如 pdf-A, .doc 等)。這份清單並非固定不變，應定期檢視與更新，以反應技術變革和組織的需求。
徵集/館藏政策	檔案館必須決定數位文件是否具有典藏價值、符合蒐集範疇。這些決定應遵循徵集或館藏政策，以確保文件是與檔案館和其使用者有關。
傳統保存政策	不論其格式為何，皆承諾提供傳統文件適當的保存，強調整體檔案館藏的保存的需求。
文件管理政策	文件管理政策協助典藏單位管理自己的數位文件，並釐清聚集和處置的方法。這些原則具有相關性，而且可能用來管理儲存已組織且聚集的數位文件。
數位永續性和資訊保證政策	數位永續政策定義資訊應如何管理，以確保資訊對典藏單位而言，仍可使用，且為完整可取用的。對於所典藏之數位文件合理地使用相同程序，使典藏單位可一致性地管理所擁有的 IT 系統。
資訊自由/資料保護政策	2000 年資訊自由法案和 1998 年資料保護法，兩者影響數位保存如何管理或傳遞。

政策	關係
	數位保存政策必須辨識典藏單位在處理這兩個法案的政策，以回應資訊需求的權威性。
資訊安全政策	為保護典藏單位的資訊財產，在資訊安全政策上將規劃特定的需求。 這些規則必須反映於數位保存政策中，以維護典藏單位的資訊安全不中斷。
著作權政策	在英國著作權法律下，無權產生文件的副本，除了某些限定的目的之外。當檔案館不是著作權擁有者時，著作權政策必須納入典藏文件的考量。
標準作業程序(SOP)	標準作業程序應提供執行數位保存任務的程序和教學，以標準方式保護數位文件的完整性。 這些標準作業程序建立成果的標竿，可與現有的成果作為對照。
服務層級協議(SLAs)	數位保存的服務層級協議應指定層級和提供支援的範圍，以管理數位保存策略中的部份或所有程序（例如：從 ICT 提供者獲得的支援管理用來儲存數位文件的伺服器）
系統發展	新系統或軟體的發展可能會造成產生新型態的數位文件。數位保存政策對早期負責數位保存者，應提供授權。 檔案館若盡早評估數位文件(例如在產生階段)，就越能準備與瞭解數位文件最佳的保存方式。
正式的合約	正式合約適當地強迫服務提供者負責，也可由數位保存政策強被作為同意合作的證據。 這個協議對中斷所同意之服務層級，應提供特定修正方式。

(六) 數位長期保存政策的評估與監控

數位長期保存政策應定期檢視政策的需求（不要超過三年），以確保政策和組織的數位保存能力，以及更大的業務目標仍是相關的。政策應成為廣大監控程序的一部分，以辨別數位保存中的任何弱點或需要改變的方法。在執行期間，政策制定後仍需持續檢視以確保其相關性與有效性。(The National Archives, 2011)

本文在檔案館制定數位保存政策之評估與監控方面，以英國國會檔案館、英國國會檔案館與赤郡檔案館及地方研究三者，列舉有關長期保存政策之評估與監控項目（見表 6），提供參考。

表 6: 長期保存政策之「評估與監控」範例

原始出處	長期保存政策之評估與監控
英國國家檔案館保存政策 (The National Archives, 2009)	英國國家檔案館的館藏維護部和數位保存部將每年重新評估保存政策。作業與服務的執行者，是策略風險擁有者和風險改善管理者，將會向國家檔案館審查委員會提供定期保證英國國家檔案館檔案文件館藏的整體性，並有效控制緩和遺失、損害。
英國國會檔案館保存政策 (Parliamentary Archives, 2009)	國會將發展與執行符合此政策的監控方式，並將執行定期稽查。這些稽查將用來評量執行政策執行的有效性，辨識未來的優先性，以及通知政策的未來評估。 此政策至少每兩年依環境變遷因素來做評估。將由國會檔案館，聯合上下議院的資訊服務部門，以及其他典藏管理者進行檢視。其他的評估將視之後的改變而定。
赤郡檔案館及地方研究的數位保存政策 (Cheshire Archives & Local studies, 2010)	最佳實務做法和技術發展在此領域變化快速，因此，政策需要每年重新做審查，修改不足地方，以因應新的問題產生。

肆、研究方法與步驟

一、研究方法

本研究以文獻分析法、深度訪談法及個案研究法進行研究問題的探討：

(一) 文獻分析法

文獻分析法是以「系統而客觀的界定、評鑑並綜合證明的方法，其主要目的在於瞭解過去、洞察現在、預測將來。」(葉至誠、葉立程，1999，p.138)

本研究欲蒐集國外數位保存相關之研究計畫、政策、後設資料、作業流程等相關書籍、期刊、論文、工具等，藉由文獻探討來了解國外的發展情況與現況，以此建立本研究的基本架構。

(二) 焦點座談法

以焦點訪談法了解中研院台史所檔案館、中研院史語所、故宮、台大數位人文中心、檔案管理局、國家圖書館等現形數位檔案保存政策、作業方式、使用的工具、檔案格式等等，透過焦點座談了解目前數位保存作業執行的現況與困境。

(三) 個案研究法

個案研究法之目的為：(1) 透過相關資料之蒐集與分析，深入瞭解所欲探討個案之問題，並找出全方位的問題解決模式；(2) 同時邀集相關人員透過溝通與協調，取得共識，並就份內工作針對問題進行適度地調整並配合之；(3) 並透過召開個案研討會，相互分享經驗，以增進其專業知能。由此可知，其目的不僅在探究問題的癥結所在，且希望能夠更具體地歸納得出適當的因應對策，並引導將解決問題的方案付諸相關的行動。(陳姿伶，2003)

本研究專注於數位保存策略的制訂與數位保存系統建置，以台史所檔案館作為深探此議題的輔佐例證之分析，焦點在於議題或事件而不在個案本身，因此屬於工具性的個案研究 (*instrumental case study*)，即個案的作用只是輔助進行議題或事件探討的工具而已。

本研究先透過深度訪談了解國內典藏單位目前數位保存的策略、實際措施、流程，並分析其實際作業問題，再參酌國外數位保存策略或標準，利用 *Archivematica* 系統作為數位保存系統原型，產出符合國際保存性後設資料標準的 XML，以訂定合適的數位保存策略與作業流程，並依據實測結果，修正數位保存策略與作業流程。

二、研究限制

本研究國內相關文獻甚少，國外機構數位保存實際案例與作業方式取得不易，

由於時間及經費限制，國外實作案例僅能以網路上可獲取之免費數位保存軟體作為實驗工具。

三、研究工具

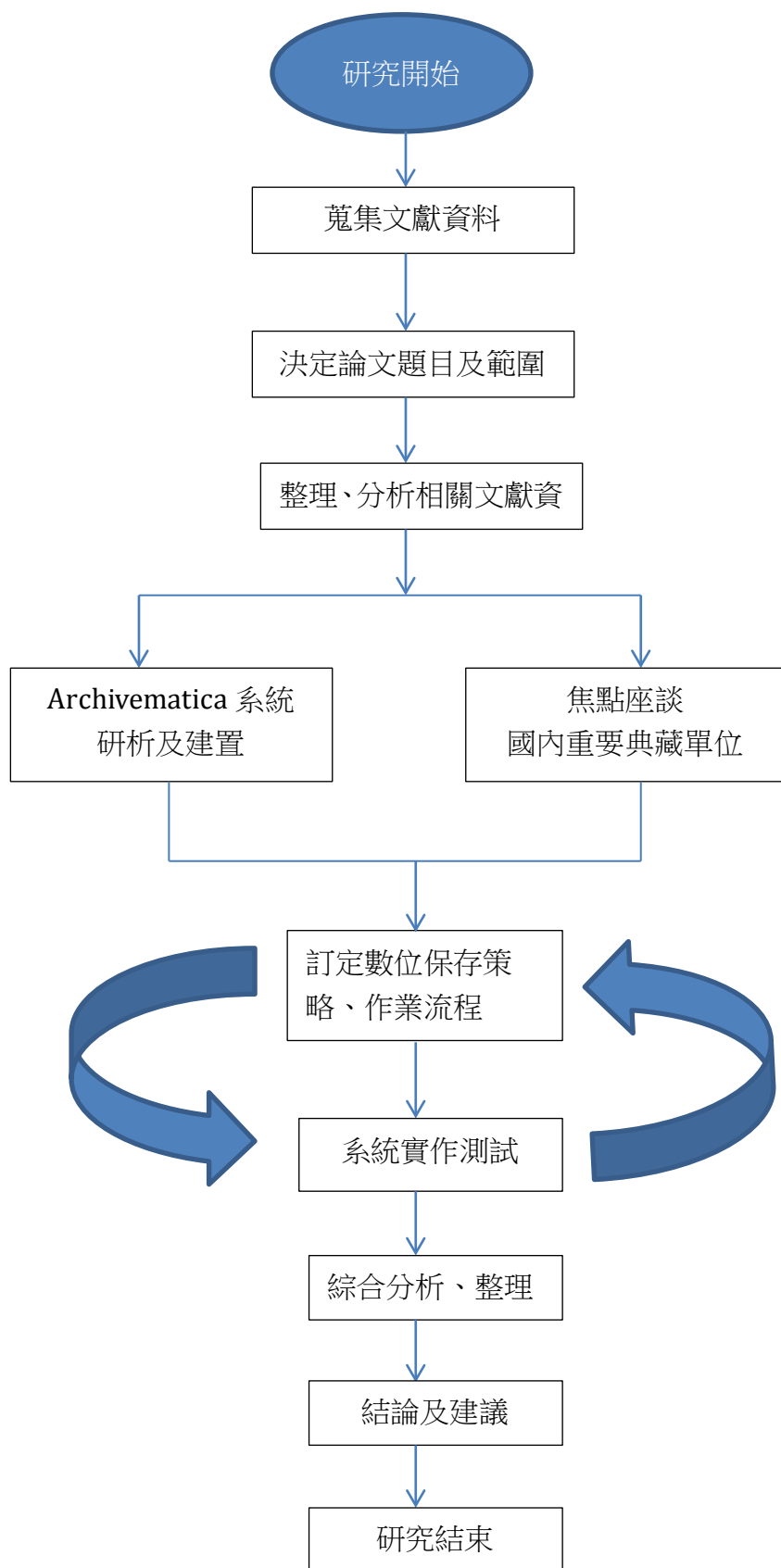


圖 2：研究程序流程圖

伍、Archivematica 軟體研析

Archivematica 為 Artefactual Systems 與聯合國世界記憶之科技委員會、加拿大溫哥華市立檔案檔、哈佛商學院貝克圖書館、紐約現代藝術博物館 (MoMA)、哥倫比亞大學圖書館、洛克斐勒檔案中心、耶魯大學圖書館、密西根大學班特列圖書館等眾多合作者共同合作發展的長期保存軟體。Archivematica 於 2013 年 5 月 1 日發佈 0.10 版軟體，至 2014 年 10 月 24 日發佈 Archivematica 1.3.0 版軟體。(Archivematica, 2014c)

Archivematica 這套數位保存系統，將數位物件攝入典藏庫作為檔案典藏，並提供典藏資料的取用，此一整體流程以自動化的方式呈現，主要是透過網路為基礎的後台管理，以圖形化介面，並協調一整套的微服務，產出符合國際標準的長期保存後資料，以作為查詢和交換分享之用。

Archivematica 整合數種開放原始碼的工具，例如 FITS、OpenOffice、FFmpeg、Clam Antivirus 使用其內部 microservices 框架，並利用開放、標準化的格式，例如 METS, PREMIS, BagIt 確認長期、標準化為基礎來管理和取用儲存於 AIP 的內容和後設資料。內容攝入 Archivematica 稱為「移轉」：包含將保存的內容，描述這些檔案的後設資料、“submission documentation” (licenses 和其他行政文件)。移轉會結構化內容，準備重新封包成為 SIP，然後轉為長期保存的 AIP，如果可提供給一般使用者使用，就會產生 DIP。

Archivematica 將上述步驟切為一系列結構化任務，在系統中稱為微服務。每一組任務，人管理人員必須做一些決策，例如正規化上傳的檔案以作為保存、取用，是否要認可正規化的結果，利用附加的描述性後資料來移轉，並儲存為 AIP。

Archivematica 的保存策略為以正規化檔案作為保存與取用格式。保存副本會加入 AIP，取用副本用來產生 DIP 上傳至取用系統。原始檔案始終都作為保存，允許在未來時進行不同的保存動作，例如正規化為不同的檔案格式或模擬。但不是所有的數位物件都可在攝入時正規化，某些數位物件例如 CAD 圖檔或 Microsoft Visio 檔案，沒有 Linux 的 opensource 工具可處理轉置，也沒有一致的保存格式。除此之外，有些格式例如 Microsoft Word 文件不必然是最好的儲存格式，且仍是普遍存在且良好支援，他們目前不需要正規化。在這些案例，Archivematica 預設是保存他們原始格式。格式風險評估和資訊由我們的使用者來取得，經過時間後，將允許人工延伸預設，透過 Format Policy Registry or FPR 包含更多格式。(Archivematica, 2014f)

雖然本計畫進行時，已有 1.0 版系統，但系統釋放不久後即發現其開放原始碼有錯誤，因此 **Artefactual systems** 暫時關閉所有分享，是故本計劃選擇 0.10 版本之系統作為測試版本。以下將就 0.10 版本之功能與流程做一介紹：

(一) 移轉

移轉是將數位物件移轉成為 SIP 的過程，Archivematica 允許以作業系統瀏覽器或網路介面進行移轉。要移轉的來源檔案，不能使用網路介面做上傳：必須要透過 Archivematica 伺服器存取。Archivematica 允許你指定“source directory”。在“source directory”裡面放置將作為移轉的檔案及目錄。一種方式為在 Dashboard 的 Transfer tab 中，使用者將數位物件透過 Storage Service 從來源目錄移動至 Archivematica。另一種方式則是在開始處理前，先建立結構化的目錄，應包含三個子目錄：logs, metadata, objects。Metadata 目錄包含 checksum, METS file, 以及 submissionDocumentation 子資料夾，可作為 transfer form, 捐贈協議或其他與徵集檔案相關的其他文件。Logs folder 包含在 Archivematica 中處理移轉時產生的 logs。在 objects 資料夾中可建立任何需要的子目錄。submission documentation 資料夾中的檔案可為捐贈協議、移轉表、著作權同意書和其他與移轉相關的文件、信件。從這次移轉的任何 SIP 封包都會自動包含這些文件的副本。在 transfer tab 中，在下拉式選單可選擇移轉類型，其包含：標準、未壓縮封包、DSpace and maildir。(Archivematica, 2014g)

當我們在準備移轉封包的內容時，封包中數位物件相對應的後設資料必須以 csv 檔案呈現，在移轉時將檔案名稱命名為 metadata.csv。Archivematica 對於 csv 檔案的規範如下：(Archivematica, 2014h)

- 第一個原始的 csv 檔案組成欄位名稱。欄位名稱不可包含 spaces
- Dublin Core 欄位名稱必須包含“dc”於名稱內，例如“dc.title”
- 對每一目錄或檔案，包含完整的欄位值
- 對多值欄位（例如 dc.subject），重複整欄，每一欄包含單一值
- 單一物件的 metadata, csv 檔案必須包含“filename”列出每一物件的檔案名稱和檔案路徑，例如：“objects/BrocktonOval.jp2”
- 複合物件的後設資料，csv 必須包含“parts”欄列出形成複合物件的包含物件（items）之目錄名稱，例如“objects/Jan021964”
- 檔案名稱可重複在其他子目錄中的檔案名稱。例如“page01.jp2”可在多個子目錄中發生

(二) 建立 SIP

確認移轉後，Archivematica 為建立 SIP 欲進行的微服務，包含 (Archivematica, 2014g)：

- 驗證符合移轉 (驗證移轉為適當的結構化—例如具有 logs, metadata and objects folders)
- 以 transfer UUID 重新命名 (指派唯一識別碼給此次移轉;移轉時會跟隨其 metadata)
- 指派 file UUIDs 以及對 object 進行 checksum (指派唯一識別碼，以及對在 /objects 目錄中的每一各檔案進行 sha-256 checksum)
- 驗證移轉的 checksum(驗證在移轉時 metadata 目錄中的任何 checksum)
- 產生 METS.xml 文件 (擷取移轉的原始順序產生 METS file。METS file 從移轉時會加入於任何產生的 SIP 封包)
- Quarantine(隔離)(隔離移轉的時間，依據 administration tab 所做的前端配置，允許在病毒掃描前更新病毒定義)
- 掃描病毒 (掃描病毒和惡意軟體)
- 淨化檔案和目錄名稱 (在資料夾和檔案名稱中移除禁止的 character 例如"&")
- 辨別檔案格式 (辨別是否需要正規化，使用者可選擇 FIDO 或於此階段跳過或延伸辨識格式)。見 Format identification
- 取得封包 (從呀所檔或其他封包檔案摘錄內容)：你可改變你預先配置的工作流程設定，允許關於封包摘錄的某些選擇。詳 Extract packages
- Characterize and 摘錄 metadata(辨別與驗證檔案格式;摘錄內嵌於檔案的技術 metadata)。如果有預先配置這些流程，Archivematica 在 micro-service 時將會停止，允許使用者從下拉式選單選擇檔案辨別的指令。預先配置選項，見 administrator manual 1.1-processing configuration。Archivematica 預設的檔案辨識，以 file extension 作為辨識。你可選擇略過辨識，之後再運作，在攝入時才運作。

移轉在處理過程中將顯示微服務已完成為綠色，若在處理中為橘色。當微服務失敗或遇到錯誤時，微服務的背景色從綠色轉為粉紅色，並在移轉旁或 SIP 名稱旁顯示"failed"。

一旦微服務完成，移轉旁會顯示鈴鐺 icon。意味著移轉已封包成為 SIP 可作為攝入或送至 backlog、索引並儲存作為日後處理時可檢索。

- Option 1：選擇”建立 SIP 並持續處理”（附註手動產生 SIP 非目前功能。從一個或多個移轉建立一個或多個 SIP 將會作為改善功能。）
- Option2：選擇”send transfer to backlog”。移轉將會於儲存 AIP 之同一地點儲存 backlog，如此就可以檢索從攝入後的一個或多個移轉。
- Option 3：選擇”reject the transfer”

（三）攝入

攝入時，數位物件會封包於 SIP 中，透過數個微服務運轉，包含正規化、封包成為 AIP，並產生 DIP。在攝入時，系統會提醒你要填入後設資料，以 Dublin Core 元素為欄位，僅能於 SIP 階段填寫後設資料。另一組的後設資料為 PREMIS Rights 欄位，但若想對封包中個別數位物件加上 Rights Metadata，則需要在物件上傳至自己的取用系統時才作。（Archivematica, 2014e）

正規化是轉置攝入的數位物件作為保存及取用格式之過程。原始物件會與正規化版本一起。正規化階段，SIP 將顯示為鈴鐺圖示，從下拉式選單選擇一種正規化：

- 正規化作為保存和取用：產生物件的保存副本，以及取用副本，將來可用來產生 DIP
 - 正規化作為取用：不產生保存副本。產生取用副本，將用來產生 DIP
 - 正規化作為保存：產生保存副本。不產生取用副本，也不會產生 DIP
 - 不作正規化：不產生保存副本。不產生取用副本，也不產生 DIP
 - 你也可能在此階段拒絕 SIP
1. 一旦正規化完成，可在正規化報告中檢視結果。點選 report icon 報告顯示什麼已經完成正規化，且什麼已是可接受的保存和取用格式
 2. 在 Review tab 可檢視正規化成果，可用瀏覽器檢視檔案，若瀏覽器無法開啟，將會下載至本地端，你可用自己電腦上適當的軟體來看。
 3. 處理 SIP 時，在 Actions 下拉式選單，認可(Approve)正規化。你也可拒絕 SIP 或再次正規化。
 - 如果在正規化中看到錯誤，在 Error handling 了解問題

（四）METS 封包

組成 AIP 的 METS file 資料目錄，以及三個資料夾：logs, objects, and thumbnails。METS file：/data/METS.uuid.xml 包含完整的 PREMIS（詳[PREMIS metadata for original file](#),[PREMIS metadata: normalized files](#),[PREMIS metadata: events](#), and[PREMIS metadata: rights](#)）（Archivematica, 2014b）

- METS file 的角色是連結原始物件至他們保存的副本，以及描述和 submission documentation，並連結 PREMIS metadata 至 AIP 中的物件
- Logs : /data/logs contains the /transfers directory，正規化 logs,惡意軟體掃描 log，以及摘錄 log（從未封包的封包）
 - /transfers目錄包含從處理到每次移轉的紀錄，此為 SIP 在移轉工作流程中的一部分
- Objects : /data/objects 包含原始物件、正規化物件，/metadata and /submissionDocumentation。如果在 SIP 終有任何低層次的目錄，目錄結構會維持
 - /metadata 包含/transfer，包含在移轉時匯入的任何後設資料
 - /submissionDocumentation 包含每次移轉時的 submission documentation，其為 SIP 的一部分，以及每次移轉的 METS.xml 檔案。移轉的結構地圖最接近移轉的原始順序
- Thumbnail:/data/thumbnails 包含在 AIP 搜尋介面檢視，所產生的任何 thumbnail

（五）儲存 AIP 或 AIC

當管理人員認可正規化後，SIP 運轉數個微服務，包含處理 submission documentation，產生 METS 檔案、索引、產生 DIP 和 AIP 封包。當這些微服務完成時，使用者可上傳 DIP 至取用系統，並儲存 AIP 封包。一旦完成攝入，在下拉式選單選擇地點以” Storing AIP” ，並壓縮與 zip AIP，並移動至檔案儲存。（Archivematica, 2014d）

因為資料集和數位物件館藏可能很大量，且為異質性，一個資料集或數位物件館藏可分成多個 AIP。這種情形時，多個 AIP 可智能地整合為一個 AIC 或 Archival Information Collection，由 OAIS 模組定義為”AIP 的內容資訊（Content Information）是其他 AIP 的聚集”。基本的 AIC 在 Archivematica 由任何數量的相關AIP,以及一個 METS 檔案包含 fileSec 以及邏輯結構圖列出所有相關的 AIP。（Archivematica, 2014a）

產生 AIC 的步驟如下（Archivematica, 2014a）：

1. 產生 AIPs
 - 1.1 攝入標準移轉的 metadata 和 data files
 - 1.2 在正規化微服務時，點選 metadata
 - 1.3 在 Metadata 之下，點選 add
 - 1.4 著錄 Dublin Core metadata。建議至少填寫 Title 以及部分的 AIC。如需要，

使用 metadata entry template 產生 AIP 內容的詳細描述

1.5 點選 Create 按鈕

1.6 回到 Ingest，完成處理並將 AIP 放入檔案儲存

1.7 重複 1.1 到 1.6，產生需要的 AIP。在 1.4 步驟確認輸入相同的 Part of AIC，如果所有的 AIP 都指定加入相同的 AIC。

1.8 一旦所有的 AIP 都放入儲存，打開 archival storage。

2. 產生 AIC

2.1 以相同值的 Part of AIC 檢索所有的 AIPs，輸入 AIC#輸入於搜尋欄，選擇 Part of AIC 作為搜尋欄。在右邊，選擇 Show AICs?，點選 Search Archival storage。螢幕會顯示搜尋結果。

2.2 產生 AIC 包含所有搜尋結果的 AIP，點選產生 AIC

2.3 在 metadata entry template，至少要填入 title 和 Identifier。Identifier 必須如 AIC 號碼的結構。如果需要時，可產生更多 AIC 內容描述

2.4 點選 Create。回到 Ingest。當 AIC 顯示時，認可 AIC

2.5 持續處理 AIC，並置放在檔案儲存

在 Archivematica 系統中，不論是 SIP、AIP、AIC 等因系統自動產生完整的 METS xml 及每一在系統所執行的微服務皆記錄於 PREMIS Metadata 中，因此皆可進行搜尋，有兩種搜尋方式，一為以 File UUID, File path, File extension, AIP UUID and AIP name 等欄位做查詢；另一種方式為依關鍵字或片語搜尋。(Archivematica, 2014d)

若需要刪除已產生的 AIP 或 AIC，則系統會詢問刪除之原因，並將刪除需求送至 Archival Storage Service 管理員，如果管理員認可需求，AIP 將可從 AIP Storage 中刪除，索引也會更新。如果管理員拒絕需求，AIP 將會保持儲存。(Archivematica, 2014d)

陸、 國內典藏單位數位長期保存管理現況

本研究為了解國內重要典藏單位進行多年數位典藏與數位學習國家型計畫後，關於長期保存數位化檔案的策略，並且是否有遭遇任何的困難。本研究先與中研院台史所檔案館進行深度訪談，以了解訪談大綱是否需要修正，之後邀請中研院史語所、國家圖書館、檔案管理局、故宮博物院、台灣大學數位人文中心進行焦點座談，此外，國史館則是以答覆問卷的方式參與本次研究。

一、 各典藏單位的原始資料與已完成數位化資料之比率

(一) 實體館藏量不斷增加，難以估算

中研院史語所、檔案管理局和國家圖書館、故宮皆因每年實體館藏資料不斷增加，使整體母數不斷擴張，而無法估計已完成數位化比率。例如，中研院史語所的内閣大庫檔案必須先經整編才能納入館藏統計，而考古文物也是每年不斷有新發現的文物，但至 102 年已產出超過 110 萬件的數位化圖檔，以及近 60 萬筆的後設資料，共計 10 個資料庫。檔案管理局每年徵集入庫的檔案數量遠比進行數位化的檔案要快且多，目前數位化比率大約為 5.71%。而國家圖書館送存的資料類別非常多，數位化比率依照各種不同資料類型而完成度不一，主要以特藏文件為優先，再以主題性方式進行數位化，另外，國家圖書館與其他典藏機構較不同的是還典藏電子論文資料、電子書、電子公報等。

(二) 以典藏數位檔案為主

中研院台史所檔案館因庫房典藏空間較小，大多和其他單位合作數位化，並共享數位化成果，因此以典藏載體以數位化檔案居多，例如與國史館台灣文獻館合作數位化臺灣總督府公文類纂、專賣局的檔案；和省諮議會合作省諮議會檔案。另外，台史所檔案館典藏的民間文書、土地契約等進行數位化，整體數位化的檔案超過台史所檔案館實體館藏的 9 成。另一特別例子為台大人文數位中心，該中心主要提供數位化技術與系統建置，原實體館藏仍典藏於台大各個系所或典藏機構。

二、 典藏數位化檔案時的考量因素

(一) 典藏數位化檔案的格式與畫素

各個典藏機構會因其母機構性質不同，對於數位化檔案格式與畫素要求相異，以中研院所屬的史語所和台史所檔案館而言，皆以學術研究為主要目的，並遵循國家型計畫所訂定的規範，文字類檔案以 300 dpi 為主，而圖像類則以 600 dpi 為主。在提供瀏覽級數位檔案時，台大數位人文中心以及國家圖書館皆為提供 150dpi 的 JPG 檔。

若為提供日後各種應用所需的檔案時，各典藏機構因應方式不一，例如中研院台史所檔案館在典藏級 TIFF 檔之外，會再轉出一份 JPG 300 的檔案，並將 JPG300 的檔案上傳至系統，並依據不同資料尺寸、類型再壓縮為瀏覽級檔案格式，因此無統一瀏覽級檔案規格。而故宮因有出版展覽目錄、輸出海報等需求，因此在最初數位化時即盡可能產生所有檔案格式，以備未來應用所需。

檔案管理局的影像規格依據文書及檔案電腦化作業規範辦理，原始檔案數位化的規格以 300dpi 為主，並分為彩色與黑白兩種格式，而線上簽核的公文則主要子常見且可在線上處理的檔案格式為主。另外，國家圖書館還典藏視聽資料，在永久保存方面以 mpg4 格式為主。

(二) 數位化檔案著錄的後設資料標準及檔案命名原則

典藏品種類較多元豐富的機構，例如中研院史語所和台大人文數位中心，其各類典藏品群會參考國際標準後，自訂後設資料欄位與著錄規範，並未採用統一的後設資料。而檔案命名原則也是相似的情形，依據各類典藏品特性訂定命名原則，例如中研院史語所各類型典藏品有各自命名原則，而台大數位人文中心亦是，並從系統端的部份進行資料的互通分享。

而以檔案為主的典藏機構，中研院台史所檔案館分有古籍系統和檔案系統，古籍系統訂有圖書書目和期刊篇目二種不同的 DTD，檔案系統則是參考 DACS 與 ISAD (G) 訂定 DTD，再轉出符合 EAD 的 XML。國史館與檔案管理局之國家檔案目錄採用 EAD 標準。在檔案命名原則方面，各機構大多依據檔案層次進行檔案編碼，因各機構檔案編排與控制層次的詳簡不一，而編碼長度也不一。

國家圖書館雖然典藏品種類眾多，但後設資料標準以 Dublin Core 為主；故宮在器物 and 書畫類方面的典藏品以 CDWA 作為為主要後設資料標準並做在地化修改，而其圖書文獻處的資料則是以 MARC 為主要的後設資料標準。此二機構在檔案命名原則方面，都有其各自的命名原則與規範。

(三) 數位典藏資料庫系統建置

各典藏機構開發之數位典藏資料庫，大多請廠商建置，在進行多年數位化工作後，某些機構開始進行系統的整合，例如中研院史語所和故宮。表 7 為各典藏機構系統建置情形。

表 7

各典藏機構系統建置方式

典藏機構	系統	備註
中研院台史所檔案館	XDCM 核心平台	
中研院史語	以 Oracle 資料庫建置，前端以	2014 年起委託大

典藏機構	系統	備註
所	JSP 做介面設計	鐸公司整合 11 個資料庫
台大人文數位中心	自行開發，也有寫 DAM	
檔案管理局	不同的原生與再生資料，使用不同的管理系統，不同的管理應用系統所對應的資料庫 dpms 就會不一樣，例如國家檔案影像處理 dpms，檔案局則是使用 microsoft 的 SQL2008r2 版本	考量原因：價格，因 dpms 是可以做變更的
國家圖書館	MAS4、MySQL、debate 等	
故宮	數位媒材管理系統式與自動化系統結合，自動化系統就可以直接點選原始的 TIFF 檔、jpg 檔	2013 年將 5 個系統進行整合，仍可提供數位媒材的 DAM
國史館	JAVA：JDK1.5、Hibernate 3.0、Spring 2.0 Apache Tomcat：5.5 以上(或同等級) 作業環境：Microsoft Windows 2008 Server(或同等級) 資料庫：Microsoft SQL Server 2008(或同等級)	

三、 數位資料的保存管理

關於數位資料的保存管理，包含數位化檔案儲存管理方式、檔案損壞與復原方式、應用軟體過時問題、格式移轉或轉置策略，分述如下：

(一) 數位化檔案儲存管理方式

中研院台史所檔案館將典藏級 TIFF 檔僅作為保存使用，並以 DVD、Hard Disk Drive 作為保存，也有部分檔案上傳至中研院計算中心。在典藏系統中的檔案都是 300dpi 的檔案，並會做備份磁帶和 Disk Arrays，而系統則會針對不同使

用需求批次轉出可查看目錄的最小縮圖，以及提供給讀者的瀏覽級圖檔。DVD、Hard Disk Drive 的保存資料內容都是一樣的，有 TIFF 檔與 JPEG 檔，在典藏系統裡面，磁帶和磁碟陣列就只有 jpg 檔，它會自動轉出瀏覽級與縮圖。但 DVD 很佔空間，都放置於庫房的櫃子上，異地備援就是電腦計算中心，或給予合作單位也持一分檔案備份。

中研院史語所數位化檔案，以硬碟和光碟為主，遵循國家型計畫的異地備份機制進行異地備份，以虛擬的磁碟陣列作為管理。主要備份於所內儲存空間（包括文物館 402 機房、傅斯年圖書館等），完成後製後，會依循國家型計畫異地備份的機制，上傳至備份空間。

台大數位人文中心，關於檔案的儲存規定至少有兩種的備份，不論是硬碟、光碟、陣列皆可，總機在台大校園裡，異地備援在水源校區，但異地備援距離是不足的。與中研院也有合作一個異地備援的機制，不過國家型計畫結束後，這部分合作也就終止了。

檔案管理局方面，數位掃描後的檔案儲存於光碟和硬碟，系統備份於磁帶，並直接備份至 NAS 系統。發生資料毀損時，以光碟和硬碟重新讀取資料；若是系統功能毀損，則是以 NAS 系統或磁帶做復原。在異地的部份，因目前影像檔過大，僅做異地備份，未做系統備援，但重要系統備援則是台北與台中同步操作。不同年代完成的光碟資料，轉入國家檔案資訊系統後，會依備援機制決定是否進行異地備援。

國家圖書館進行數位化檔案掃描時，會儲存於 DVD、Hard Disk，一份備份儲存於網路上，會定期備份至 Hard Disk，但有的存放於磁碟陣列。目前異地備份位於資訊圖書館，一部分位於書庫。異地備援位於宜蘭大學，但僅網頁電子書送存系統、ISDN 系統、館藏系統有做異地備援。

故宮的數位化檔案備份總共有 4 份，主要儲存於磁碟、磁帶（原為 LTO2，現更換為 LTO5）、DVD（DVD 於 2013 年都轉置為藍光，燒錄第二份的 LTO 帶），在故宮南院進行異地備份，待故宮南院興建完成後，會再進行異地備援。基本的檔案備份，皆是磁帶再加一份備份。

（二）數位化檔案損壞之復原方式

在焦點座談中，各典藏機構提到多種數位化檔案損壞的情形，本研究將之歸

為四類，如下所述：

1. 磁碟壞軌或磁碟陣列毀損

有些圖檔在進行掃描數位化時，即已發生壞軌，導致圖檔模糊，此種方式需要重新進行掃描。若是陣列損毀或硬碟損毀，則是利用備份迅速重建資料。

2. 光碟損毀

關於離線儲存的媒體，例如光碟片，因保存期限難以判斷，各典藏機構當需使用光碟中的資料時，偶而會發生檔案無法開啟的問題，此時只能找出另一備份檔案使用，若同樣無法讀取，就必須重新掃描檔案。檔案管理局每年定期清查離線儲存媒體的讀取狀況，針對讀取有問題的影像內容，會直接利用另一套儲存媒體進行複製，必要時再調原件重新數位化。此外，檔案管理局也已將 VCD 皆以轉置為藍光儲存。

3. 軟體損壞

有時數位典藏系統會發生損壞或癱瘓的問題，但因最初開發資料庫的單位，沒有留下程式說明及系統損壞的應變方式，導致典藏單位最後必須將應用軟體重新轉置，建立新的系統作為保存與服務。

另外，數位典藏系統有時發生檔案連結錯誤，此類多是在移機時發生或系統中的連結失效。

4. 駭客入侵

有時數位典藏系統會因駭客惡意破壞，使得系統無法運作，此時也是以另一備份做迅速的復原。檔案管理局表示目前政府機關處理關鍵基礎建設部分，有要求要做安全監控，現在分成 2 個層級，以前是知通安全彙報，針對安全基礎建設設置一個設備回傳資料，現在則是讓民間的廠商去做，例如中華電信、宏碁等，和他們取得監控的機制，再做安全彙報，然後再彙整全國性的資料，再決定哪些資料需要做防毒，目前是用這種背景的資料庫來進行的。

(三) 應用軟體過時

由於數位化的檔案影像多以 TIFF 和 JPEG 格式儲存，目前尚未發生檔案

格式過時的問題。但應用軟體或系統支援上，則有過時或無法支援的問題。例如，中研院史語所傅斯年圖書館從1988年起開始進行善本古籍的數位化工作，選擇了當時最先進的「王安整合影像資訊系統」，利用光碟儲存館內善本圖籍，並建立書目與影像連結之資料庫。但1996年王安公司不再支援該系統，因而需進行系統轉換。現在該館則用非專屬化的格式和系統，避免問題再發生。檔案管理局方面，則有公文交換標準格式因廠商倒閉，而使得必須撰寫其他保存格式的程式，來保存電子公文。故宮則曾發生LTO磁帶過時的問題，原LTO1的閱讀機器和燒錄的軟體版本、以及當時的作業系統都已找不到，復原方式則將所有的資料全部重新讀取一次，並重新轉置。

（四）格式移轉或轉置策略

台大數位人文中心的格式轉置，為二年一次，利用系統寫程式提醒須轉置的時間，但有時因需要轉置的資料太多，並未確實執行轉置工作。國家圖書館則是將啟動長期保存計畫，欲將所有儲存於光碟中的數位資料全部讀取一次，並更新為磁帶儲存，希望將數位物件和後設資料可以先做一個封裝，再以檔案的格式去做儲存，希望可以每三年做一次更新，但此計畫仍未開始。

關於原生性的軟體，檔案管理局自己開發保存的工具以及共用的元件，提供機關下載使用。因電子公文沒有實體資料，保存上較為複雜，必須以公文樣態保存，以封裝的格式來提供，檔案局提供如何開啟此封裝檔的內容，為了確保它的內容不會被變更，所以在封裝的時候，必須加上憑證加簽的動作。如果XP系統無法使用，XP系統的檔案在新的系統上要讀取，就要用模擬的方式，只能用相容方式，較常使用轉置成現在使用的格式，這種方式就常被使用，轉置的工作很費時，所以通常用系統去幫忙做輔助。

柒、Archivematica 實測— 以台史所檔案館典藏品為例

本研究在了解國內典藏單位進行數位化檔案長期保存方式與策略後，並建置Archivematica 0.10版系統，以中研院台史所吉岡喜三郎文書全宗作為測試實例。吉岡喜三郎文書全宗是日籍人士吉岡喜三郎(1882-196?)所產生的文書，時間橫跨明治40年至昭和14年間(1907-1939)，絕大部分是其於日治時期來臺擔任警察工作期間(1907-1931)產生的。文書的內容包含吉岡喜三郎的私人日記、工作相關的各式書狀、履歷文件、照片等。本批文書最早是由戰前臺灣文物收藏家鄭喬維先生及國立臺灣師範大學歷史所曾令毅先生所發掘。基於長久保存臺灣珍貴史料的想法，文書發現者與吉岡喜三郎家屬(文書持有者)持續洽商，爭取將文書留存於臺灣。經過將近一年的溝通，使吉岡家了解本批文書對於臺灣歷史研究的

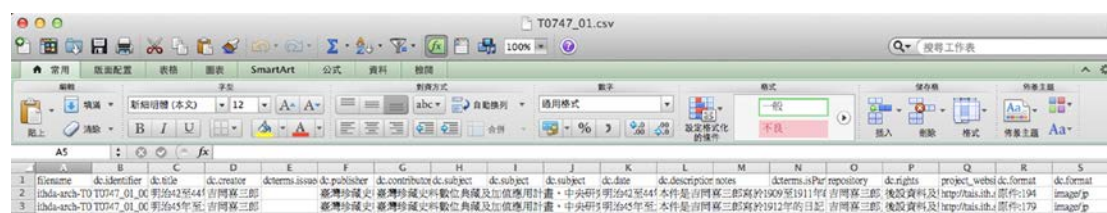
重要性與特殊性，更加重視史料的保存與利用，因此認同將文書交由臺灣公家學術機關保存的提議。透過曾令毅先生的中介，吉岡家屬於 2010 年 3 月訪問中研院臺史所，並捐贈吉岡喜三郎文書。本全宗共分三次入藏，前二批是由家屬在 2010 年 3 月、5 月親自送到臺史所，包含日記、照片、書狀及履歷文書；第三批是是吉岡喜三郎明治時期日記，由臺史所「臺灣文史資源海外徵集與國際合作計畫」（數位典藏國家型計畫）團隊，於同年 6 月從日本取回。（中央研究院臺灣史研究所臺灣史檔案資源系，無日期）

一、建立移轉封包與 SIP

在本次測試資料中，主要為三個系列：吉岡喜三郎文書 日記(T0747_01)、吉岡喜三郎公職關係文書(T0747_02)、吉岡喜三郎在臺時期照片(T0747_03)。準備移轉至 Achivematica 系統的封包，主要以一個系列的資料做為一個封包，封包包含數位圖檔及 metadata csv 檔，資料夾以系列號命名(實際範例詳圖 3)。首先需將原數位典藏系統匯出之 Metadata XML 轉為符合保存系統所需之 metadata.csv 檔，保存系統將會解析.csv 轉為 METS XML。

圖 3

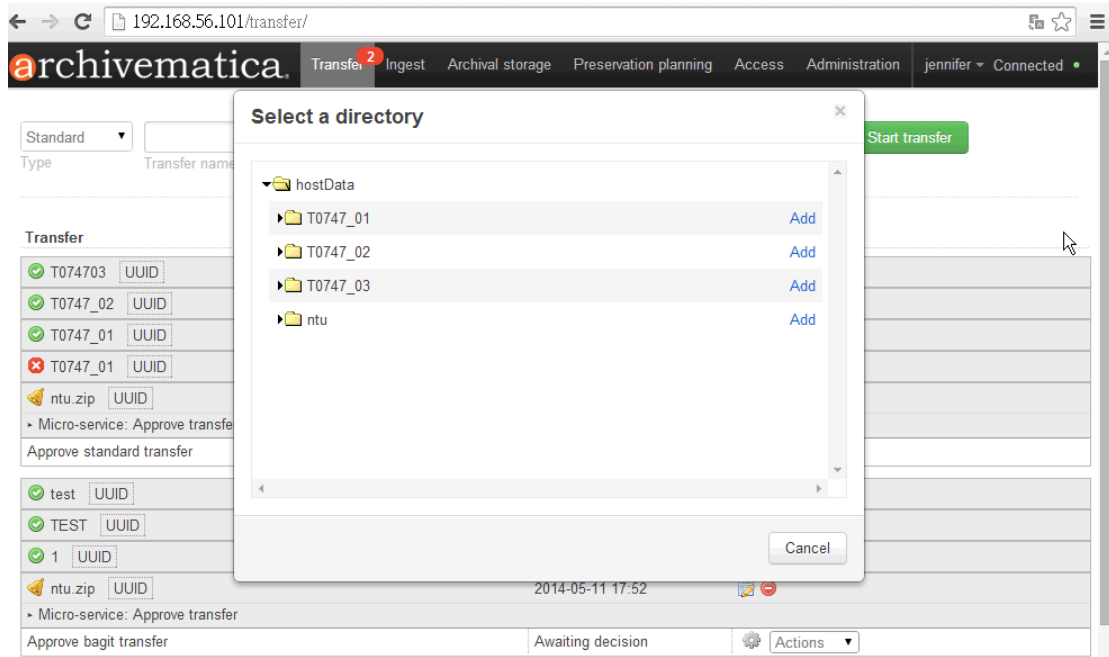
符合保存系統規範之.csv 檔範例



第二步驟為將欲上傳至保存系統的完整封包（包含數位圖檔及相對應 csv 檔），上傳至虛擬機器中的保存系統資料夾，之後登入保存系統，在 Transfer 頁面下選擇預上傳的封包資料夾，由於計畫時間有限，我們選擇以系統預設的標準模式進行典藏保存。

圖 4

選擇預上傳資料夾，並填入登錄號，開始上傳



系統進行上傳後，會請系統操作者確認此封包是否要移轉或拒絕移轉。若確認移轉，則系統會開始進行一系列的微服務，詳見圖 5 與圖 6。

圖 5

保存系統要求確認是否移轉

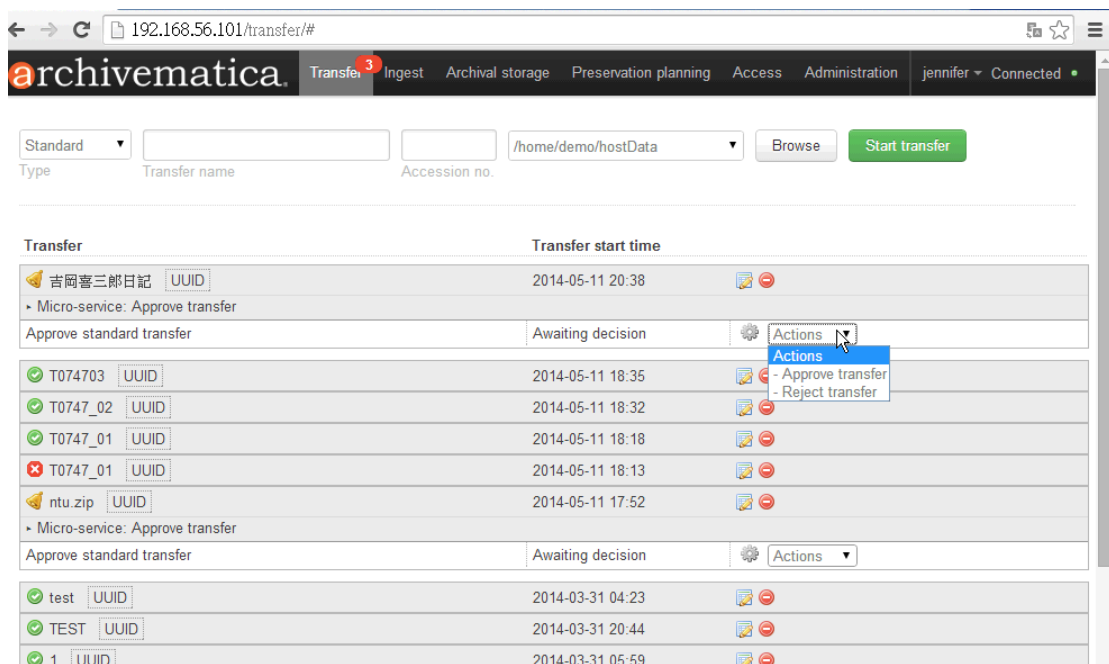
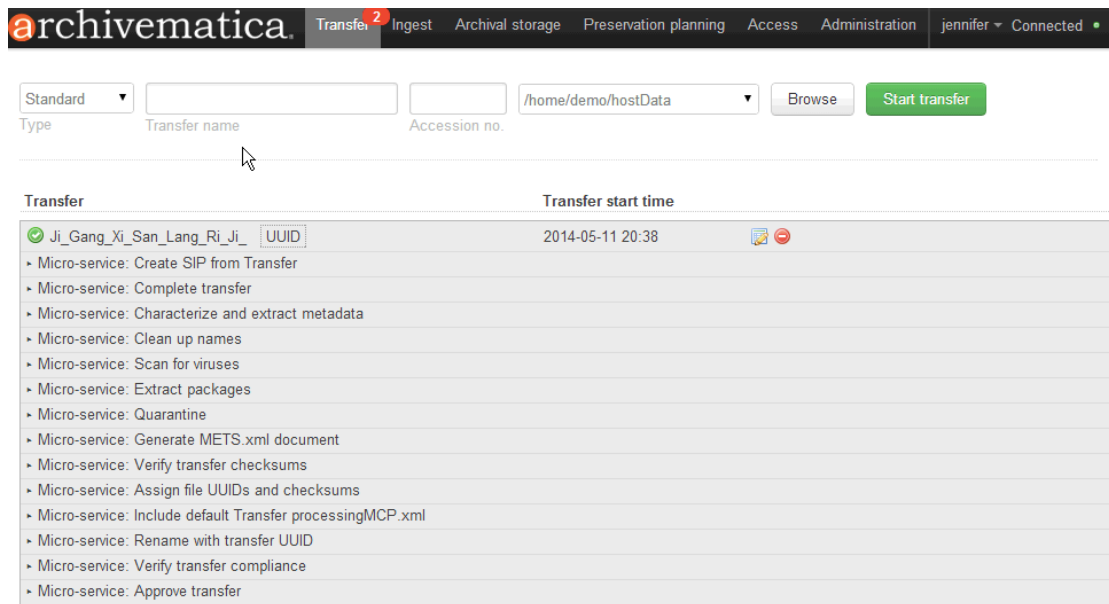


圖 6

系統自動進行微服務任務



微服務完成後，系統會告知已成功完成微服務，並可查閱微服務執行報告，詳圖 7 和圖 8。此時 SIP 封包已完成，並建立 AIP 封包。

圖 7

微服務執行的細項內容

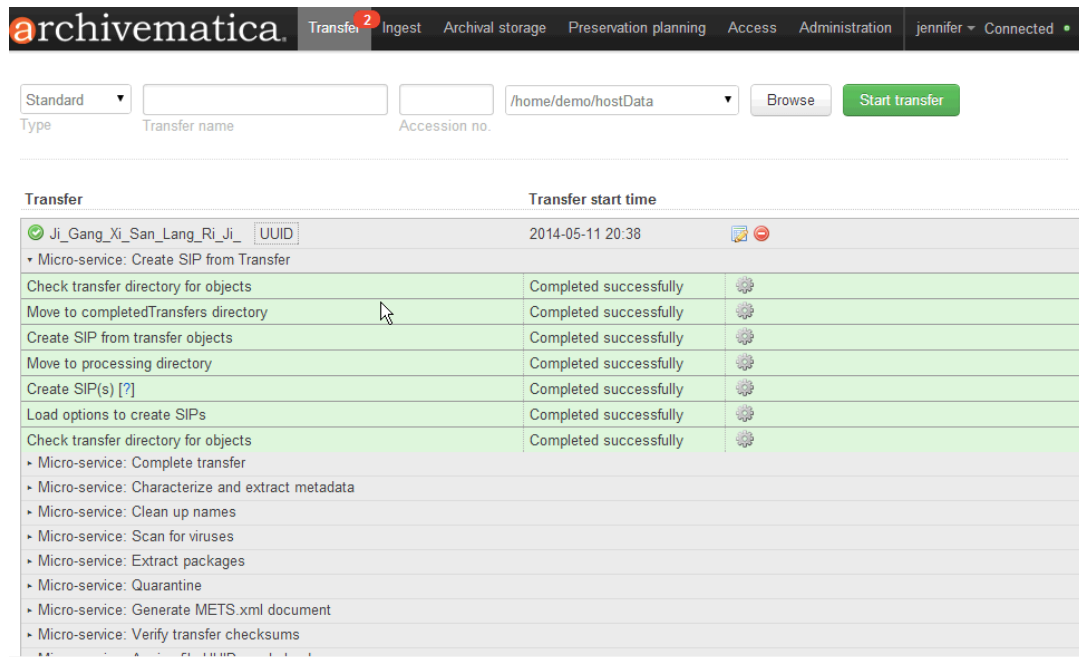
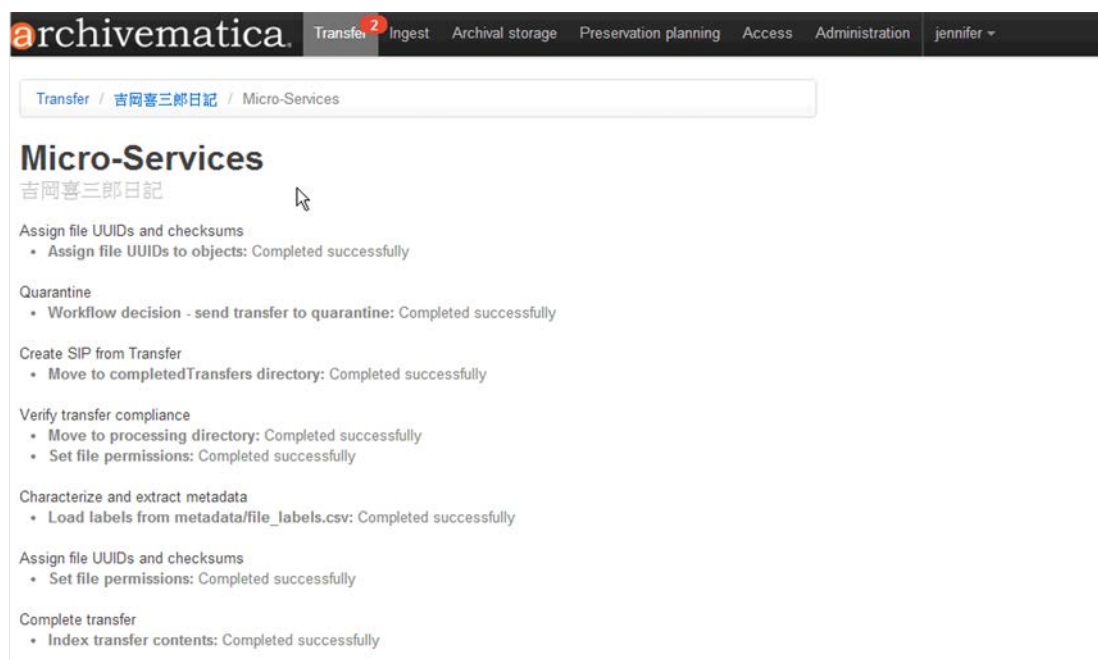


圖 8

微服務執行報告



若在移轉封包中，未包含後設資料檔案，則在攝入時，系統會提醒必須填入描述性或權利的後設資料，此時可於系統中手動填入後設資料，值得注意的是目前該系統無法接受中文輸入。權利性後設資料也必須按照檔案館實際作業與政策來填寫，詳圖 9 與圖 10。

圖 9

補填後設資料



圖 10

權利性後設資料表單

The screenshot shows the 'Rights' form in Archivematica. The form is titled 'Rights test02'. It includes a 'Basis' dropdown menu set to 'Copyright'. Below this are several text input fields: 'Copyright status', 'Copyright jurisdiction' (with a tooltip that reads 'a coded designation of the copyright status of the object at the time the rights statement is recorded; eg copyrighted, publicdomain, unknown'), 'Copyright determination date', 'Copyright start date', and 'Copyright end date'. Each date field has a note below it: 'Use ISO 8061 (YYYY-MM-DD)'.

二、正規化服務

以數位化圖檔為例，Archivematica 系統以 TIFF 檔作為保存格式，若今日在移轉封包中準備 JPG 檔案格式的數位化圖檔，必須依照館方政策選擇適當的正規化策略，完成後可檢視正規化的結果，包含原始上傳圖檔、正規化後的保存格式圖檔以及瀏覽級的 thumbnail，另外則是產出正規化程序的相應 XML。管理人員也可確認/拒絕/重作正規化動作。

圖 11

正規化策略的選擇

The screenshot shows the Archivematica interface with a search bar at the top. Below the search bar, there is a table with columns for 'Submission Information Package' and 'Ingest start time'. The table contains a list of tasks for a submission named 'test02'. The 'Actions' menu is open for the first task, 'Normalize [?]', showing options like 'Normalize for preservation and access', 'Normalize for preservation', 'Reject SIP', 'Normalize service files for access', 'Do not normalize', 'Normalize manually', and 'Normalize for access'.

Submission Information Package	Ingest start time
test02 [UUID]	2014-01-16 23:43
• Micro-service: Normalize	
Normalize [?]	Awaiting decision
Resume after normalization file identification tool selected.	Completed successfully
Set SIP to normalize with file extension file identification.	Completed successfully
Select format identification tool	Completed successfully
Set resume link after tool selected.	Completed successfully
Find options to normalize as	Completed successfully
Move to workFlowDecisions-createDip directory	Completed successfully
Grant normalization options for no pre-existing DIP	Completed successfully
Set remove preservation and access normalized files to renormalize link.	Completed successfully
Check for Access directory	Completed successfully
Check for Service directory	Completed successfully
Identify manually normalized files	Completed successfully
• Micro-service: Clean up names	

圖 12
檢視正規化結果

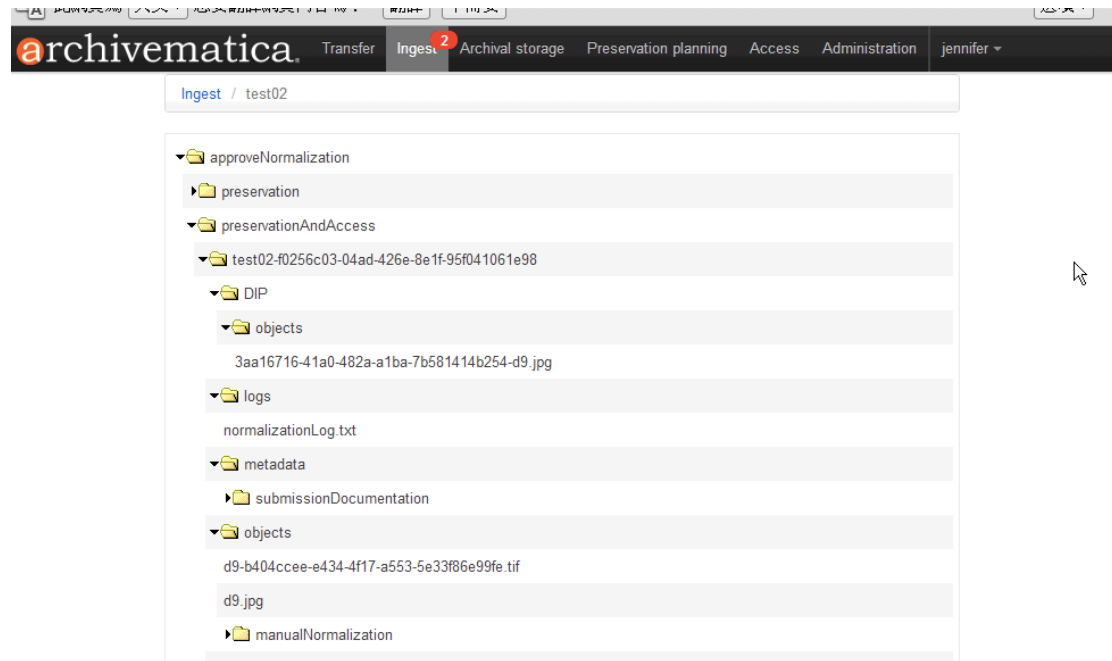


圖 13
檢視原始圖檔

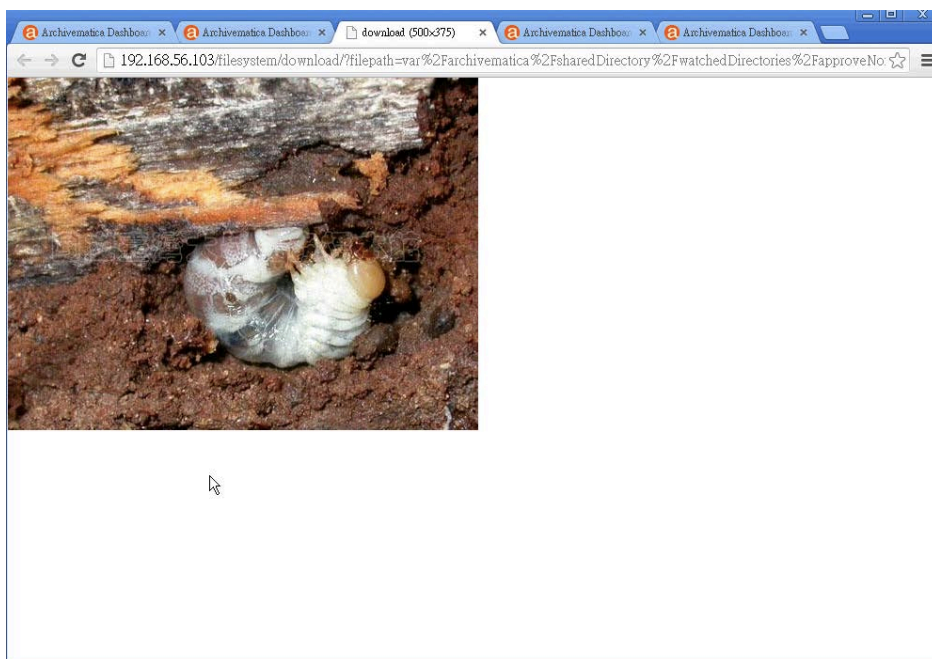


圖 14
檢視正規化後的圖檔

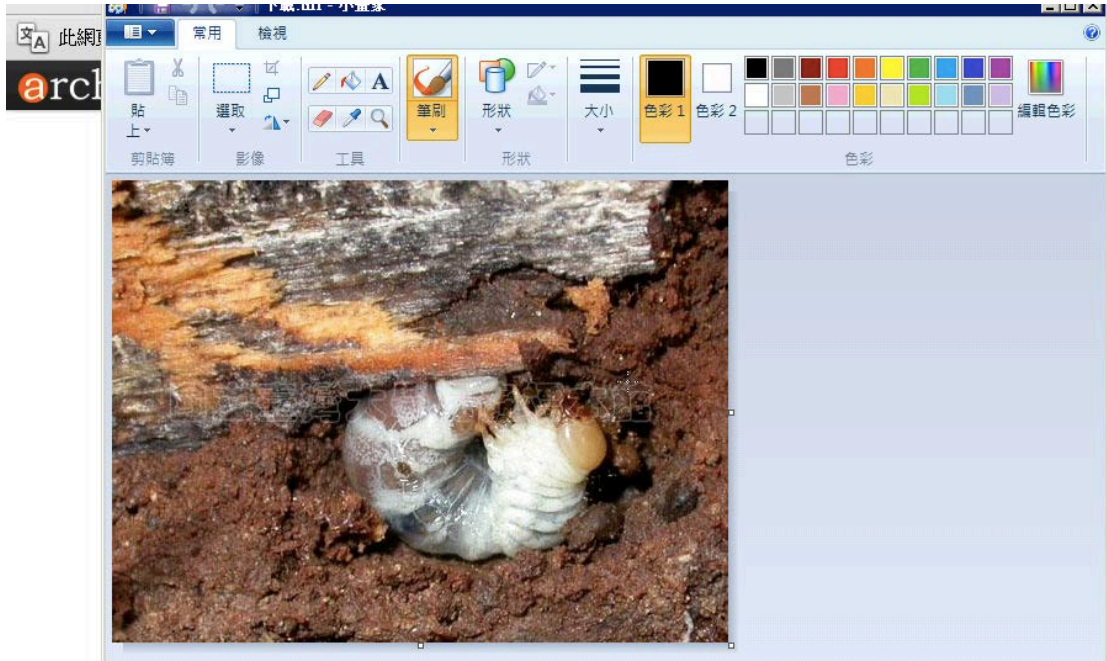


圖 15
檢視 thumbnail



圖 16
因正規化程序所產生的 xml

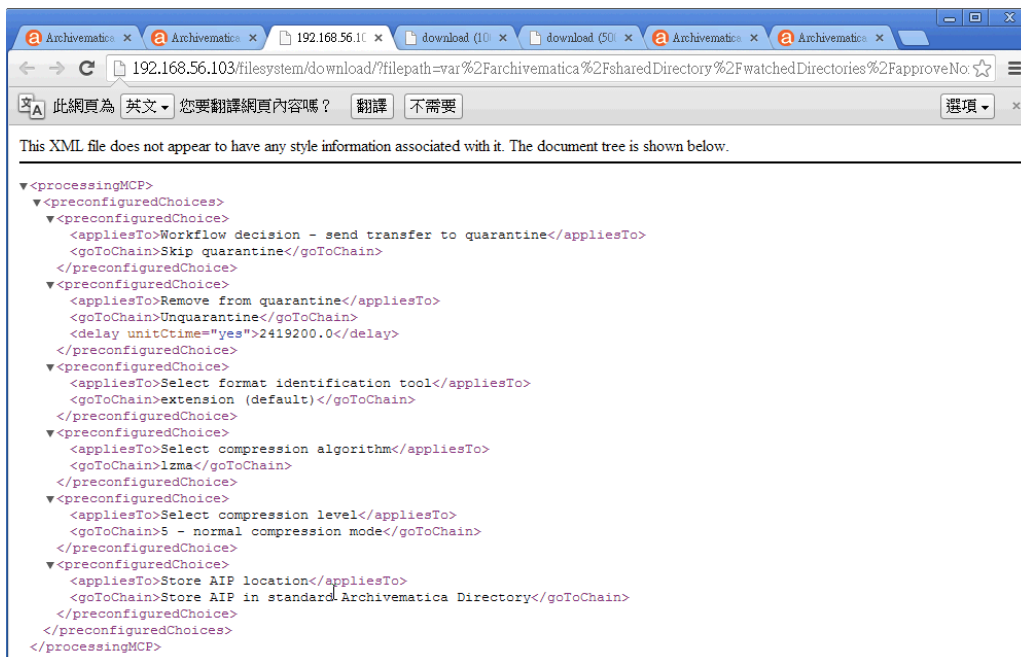


圖 17
正規化的確認/拒絕/重做

Any Keyword Show files?

Submission Information Package Ingest start time

test02	UUID	2014-01-16 23:43		
Micro-service: Normalize				
Approve normalization (review) [?]	Awaiting decision		Actions	
Move to approve normalization directory	Completed successfully		- Reject	
Remove files without linking information (failed normalization artifacts etc.)	Completed successfully		- Approve	
Set file permissions	Completed successfully		- Redo	
Normalize for preservation and access	Completed successfully			
Create thumbnails directory	Completed successfully			
Create DIP directory	Completed successfully			
Move to processing directory	Completed successfully			
Normalize [?]	Completed successfully			
Resume after normalization file identification tool selected.	Completed successfully			
Set SIP to normalize with file extension file identification.	Completed successfully			
Select format identification tool	Completed successfully			
Set resume link after tool selected.	Completed successfully			
Find options to normalize as	Completed successfully			
Move to workFlowDecisions-createDip directory	Completed successfully			

三、AIP 封包

在正規化完成後，產生典藏的 AIP 封包，在保存系統可瀏覽、查詢以儲存的 AIP 封包，封包中包含組成 AIP 的 METS file 的資料目錄，以及三個資料夾：logs, objects, and thumbnails。

圖 18

瀏覽保存系統中已典藏的 AIP 封包

192.168.56.101/archival-storage/

archivematica Transfer **Ingest** 2 Archival storage Preservation planning Access Administration jennifer

Any Keyword Show files?

Add New

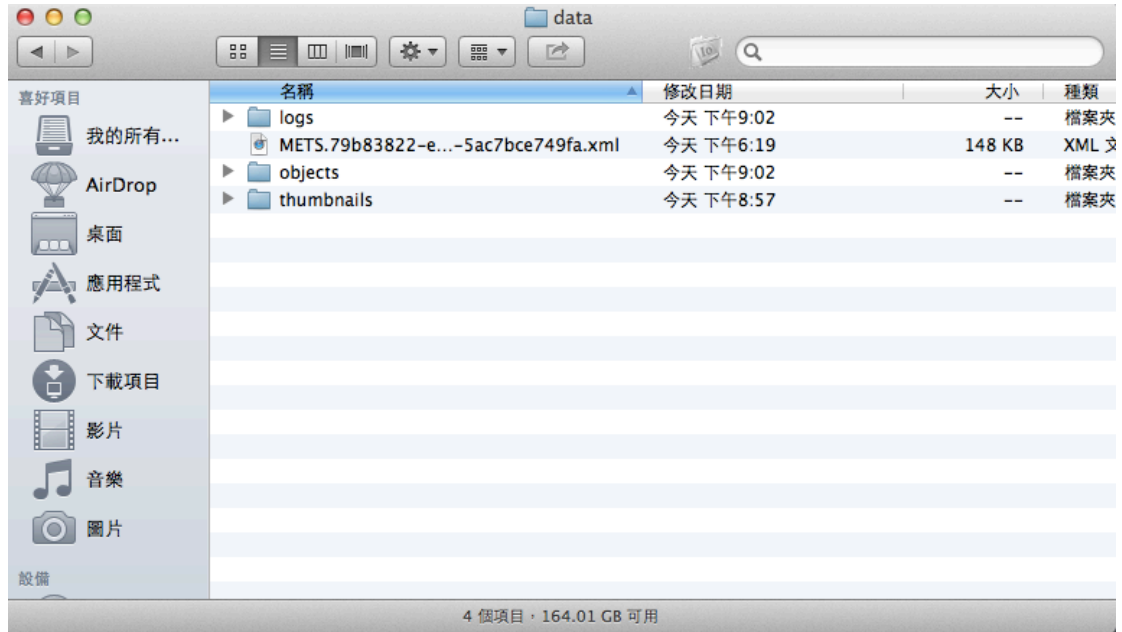
Browse archival storage

Total size: 5.04 MB Total files: 34 indexed

AIP	Size	UUID	Date stored
1	0.50 MB	29e10453-9ee8-4d9e-9c46-339d81c274f8	2014-03-31 06:00
123test	0.50 MB	e425d736-ca7b-410e-901d-75543c028412	2014-03-31 05:04
final	0.50 MB	03899172-fba4-4e84-a8ff-585bab21d2cb	2014-03-31 04:56
Ji_Gang_Xi_San_Lang_Ri_Ji_	0.49 MB	bfe4800d-6201-4ce1-803a-0ea5d31f0cee	2014-05-11 20:46
T074703	0.47 MB	69155e88-0930-4124-b4fd-08cbe30d3d17	2014-05-11 18:37
T0747_01	0.49 MB	79b83822-ec41-4e35-b250-5ac7bce749fa	2014-05-11 18:19
T0747_02	0.59 MB	e55746fd-6e21-4f83-97dc-443bf5a55145	2014-05-11 18:34
test	0.50 MB	08e7d355-51ea-4798-954c-8beab4a9ef5d	2014-03-31 04:25

圖 19

AIP 封包內容



四、DIP 封包

Archivematica 會在我們移轉封包後，自動進行一連串的处理工作，最後產生 DIP 並且儲存在 Archivematica 系統裡，資料夾路徑為：

`/var/archivematica/sharedDirectory/watchedDirectories/uploadedDIPs/`

除此之外，還可以將 DIP 傳送到 AtoM 資料庫或 Dspace 等平台或雲端系統作備份，並且有 GUI 平台供使用者搜尋與管理資料。

四、符合 PREMIS 標準的 XML

Archivematica 軟體在進行一連串微服務時，將所有執行任務的代理者、軟體、執行的任務與結果，都以 PREMIS 保存性後設資料記錄下來，現以 JPEG 圖檔進行正規化為 TIFF 檔作為實例說明。

圖 20

Archivematica 系統攝入 JPEG 檔案之 PREMIS 模組示意圖

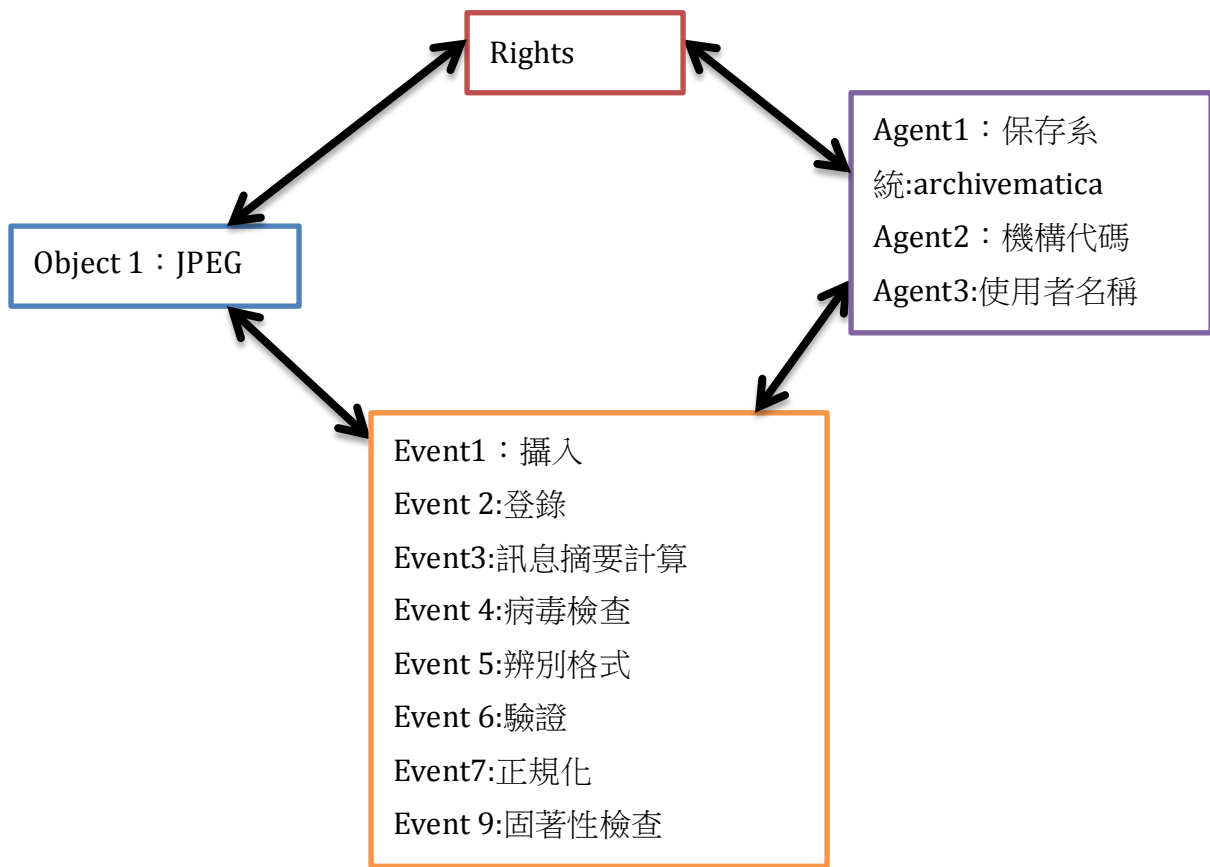


圖 21

Archivemata 對所攝入檔案所產生之 PREMIS XML

```

- <amdSec ID="amdSec_1">
- <techMD ID="techMD_1">
- <mdWrap MDTYPE="PREMIS:OBJECT">
- <xmlData>
- <object xsi:type="file" xsi:schemaLocation="info:lc/xmlns/premis-v2 http://www.loc.gov/standards/premis/v2/premis-v2-2.xsd"
version="2.2">
- <objectIdentifier>
<objectIdentifierType>UUID</objectIdentifierType>
<objectIdentifierValue>b85ce231-fc8f-4ee0-90a6-e222a6215f91</objectIdentifierValue>
</objectIdentifier>
- <objectCharacteristics>
<compositionLevel>0</compositionLevel>
- <fixity>
<messageDigestAlgorithm>sha256</messageDigestAlgorithm>
- <messageDigest>
7f8eaacfb40ee64308487ade87ebe1ebceb88c66789d64bd2f415e58355f3e9b
</messageDigest>
</fixity>
<size>1486</size>
- <format>
- <formatDesignation>
<formatName>Extensible Markup Language</formatName>
<formatVersion>1.0</formatVersion>
</formatDesignation>
- <formatRegistry>
<formatRegistryName>PRONOM</formatRegistryName>
<formatRegistryKey>fmt/101</formatRegistryKey>
</formatRegistry>
  
```

圖 22

Archivematica 對攝入事件產生的 PREMIS XML

```
--<mdWrap MDTYPE="PREMIS:EVENT">
--<xmlData>
--<event xsi:schemaLocation="info:lc/xmlns/premis-v2 http://www.loc.gov/standards/premis/v2/premis-v2-2.xsd" version="2.2">
--<eventIdentifier>
--<eventIdentifierType>UUID</eventIdentifierType>
--<eventIdentifierValue>359655dd-60a4-424f-946d-00c073dc9fd3</eventIdentifierValue>
</eventIdentifier>
<eventType>ingestion</eventType>
<eventDateTime>2014-05-11T10:19:28</eventDateTime>
<eventDetail/>
--<eventOutcomeInformation>
--<eventOutcome/>
--<eventOutcomeDetail>
--<eventOutcomeDetailNote/>
</eventOutcomeDetail>
</eventOutcomeInformation>
--<linkingAgentIdentifier>
--<linkingAgentIdentifierType>preservation system</linkingAgentIdentifierType>
--<linkingAgentIdentifierValue>Archivematica-0.10</linkingAgentIdentifierValue>
</linkingAgentIdentifier>
--<linkingAgentIdentifier>
--<linkingAgentIdentifierType>repository code</linkingAgentIdentifierType>
--<linkingAgentIdentifierValue>nccu01</linkingAgentIdentifierValue>
</linkingAgentIdentifier>
</event>
</xmlData>
</mdWrap>
--</pre>
```

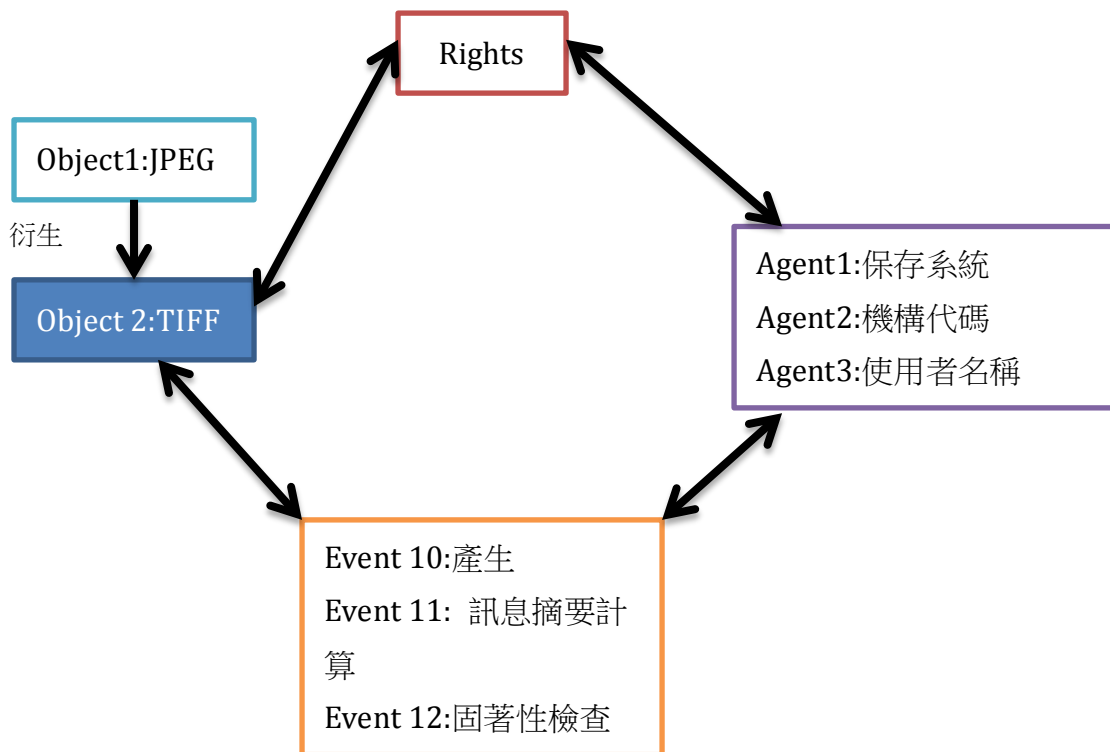
圖 23

Archivematica 對代理者所產生的 PREMIS XML

```
--<digiprovMD ID="digiprovMD_7">
--<mdWrap MDTYPE="PREMIS:AGENT">
--<xmlData>
--<agent>
--<agentIdentifier>
--<agentIdentifierType>preservation system</agentIdentifierType>
--<agentIdentifierValue>Archivematica-0.10</agentIdentifierValue>
</agentIdentifier>
--<agentName>Archivematica</agentName>
--<agentType>software</agentType>
</agent>
</xmlData>
</mdWrap>
</digiprovMD>
```

圖 24

Archivematica 系統正規化 TIFF 檔案之 PREMIS 模組示意圖



捌、結論

檔案館若採用 Archivematica 作為長期保存軟體，即使僅使用系統預設功能，檔案人員也能完成數位保存管理的任務。Archivematica 能指派每一物件、事件、代理者不同的識別碼，並能將數位物件和其對應之描述性後設資料一起封存，能解決過去數典計畫將數位檔案與其後設資料分開存放的問題。Archivematica 使用 METS 和 PREMIS 標準記錄數位保存程序、工具與日期，使得檔案人員可更有效的管理數位保存。

參考書目

中央研究院臺灣史研究所臺灣史檔案資源系統(無日期)。吉剛喜三郎文書全宗。

檢自：

<http://tais.ith.sinica.edu.tw/sinicafrsFront/browsingLevel1.jsp?xmlId=0000251776>

中央研究院臺灣史研究所檔案館。(2012)。關於本館

<http://archives.ith.sinica.edu.tw/zh-about> (2012/12/18)

王文英(2003)。中研院近史所檔案館數位保存後設資料項目建置之研究。國立政治大學圖書資訊與檔案學研究所。未出版碩士論文。

林巧敏、鄭伊廷(2011, 6月)。臺灣地區檔案典藏機構數位化作業現況分析。檔案季刊, 10(2), 76-93。

陳姿伶(2003)。個案研究法(Case Study)

http://www.extension.org.tw/book/02_92-1.9.doc (2012/12/21)

黃文彥、林信成(2012)。基於 OAIS 之開放取用期刊管理系統設計與實作。教育資料與圖書館學, 49(4), 507-532。

Archivematica(2014f). AIC. Retrieved from:

<https://www.archivematica.org/wiki/AIC>

Archivematica(2014h).AIP structure. Retrieved from:

https://www.archivematica.org/wiki/AIP_structure

Archivematica.(2014a).Main Page. Retrieved from:

https://www.archivematica.org/wiki/Main_Page

Archivematica(2014e).UM archival storage 1.1. Retrieved from:

https://www.archivematica.org/wiki/UM_archival_storage_1.1#Storing_the_AIP

Archivematica.(2014c).UM ingest 1.1. Retrieved from:

https://www.archivematica.org/wiki/UM_ingest_1.1

Archivematica(2014g). UM preservation planning 1.1. Retrieved from:

https://www.archivematica.org/wiki/UM_preservation_planning_1.1

Archivematica.(2014b).UM transfer 1.1.Retrieved from:

https://www.archivematica.org/wiki/UM_transfer_1.1#Create_a_transfer

Archivematica(2014d).UM Transfer metadata import 1.1. Retrieved from:

https://www.archivematica.org/wiki/UM_Transfer_metadata_import_1.1

Cheshire Archives & Local studies.(2010). Digital preservation policy.

Retrieved from:

http://archives.cheshire.gov.uk/record_care/digital_preservation/digital_preservation_policy.aspx

National Archives of Australia.(2011).Digital preservation policy. Retrieved from:
<http://www.naa.gov.au/about-us/organisation/accountability/operations-and-preservation/digital-preservation-policy.aspx>

OCLC/RLG Working Group on Preservation Metadata: Implementation Strategies (PREMIS) . Data dictionary for preservation metadata: final report of the PREMIS Working Group (OCLC/RLG:2005).
http://www.loc.gov/standards/premis/v1/premis-dd_1.0_2005_May.pdf(2012/12/18)

Parliamentary Archives.(2009).A digital preservation policy for parliament. Retrieved from:
<http://www.parliament.uk/documents/upload/digitalpreservationpolicy1.0.pdf>

PREMIS Editorial Committee(2012)PREMIS Data Dictionary for preservation metadata version 2.2
<http://www.loc.gov/standards/premis/v2/premis-2-2.pdf> (2012/12/18)

Sheldon, Madeline. (2013). Analysis of current digital preservation policies – archives, libraries, and museums. Retrieved from:
<http://www.digitalpreservation.gov/documents/Analysis%20of%20Current%20Digital%20Preservation%20Policies.pdf?loclr=blogsig>

The National Archives.(2009).Preservation policy. Retrieved from:
<http://www.nationalarchives.gov.uk/documents/tna-corporate-preservation-policy-2009-website-version.pdf>

The National Archives.(2011).Digital preservation policies: guidance for archives. Retrieved from:
<http://www.nationalarchives.gov.uk/documents/information-management/digital-preservation-policies-guidance-draft-v4.2.pdf>

附錄：Archivematica 系統實作

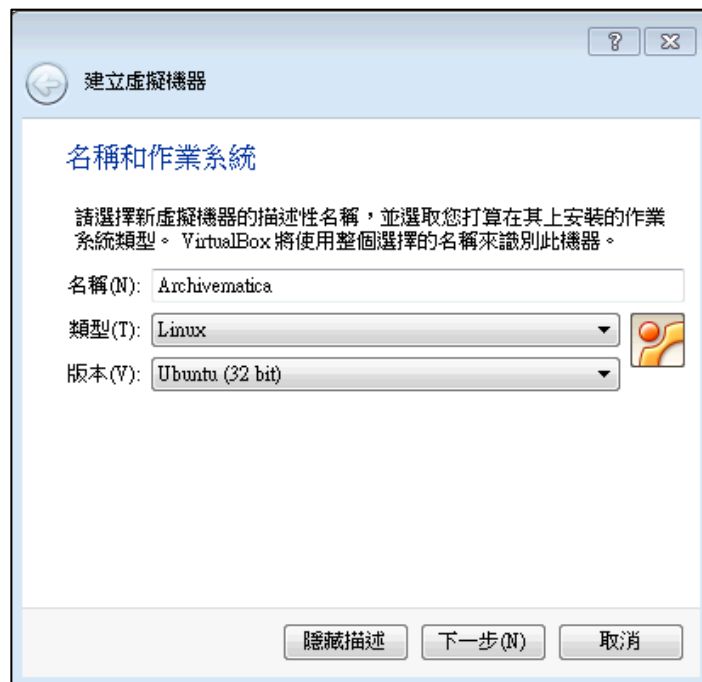
我們使用的 Archivematica 版本為 0.10-beta 版本，雖然系統建構當時已釋放 1.0 版本，但由於在釋放沒多久後發現其開放原始碼有錯誤，因此 Artefactual systems 暫時關閉所有分享，因而我們選擇 0.10 版本之系統。此小節將說明系統安裝流程，使用者帳號以 demo 為例。

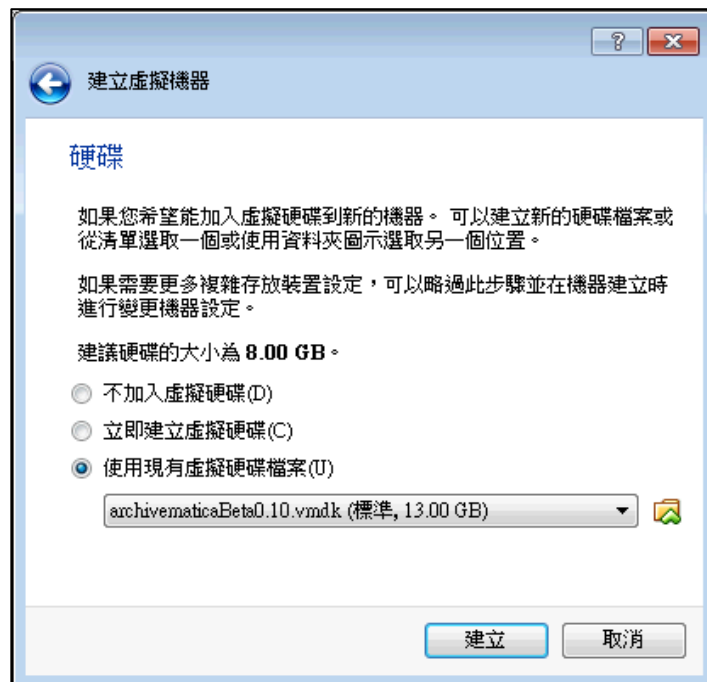
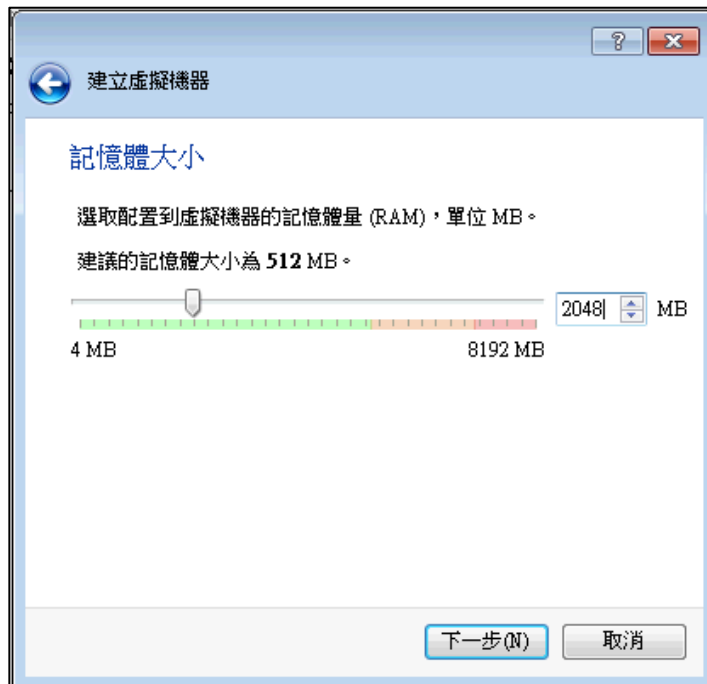
一、 建置 Archivematica 伺服器

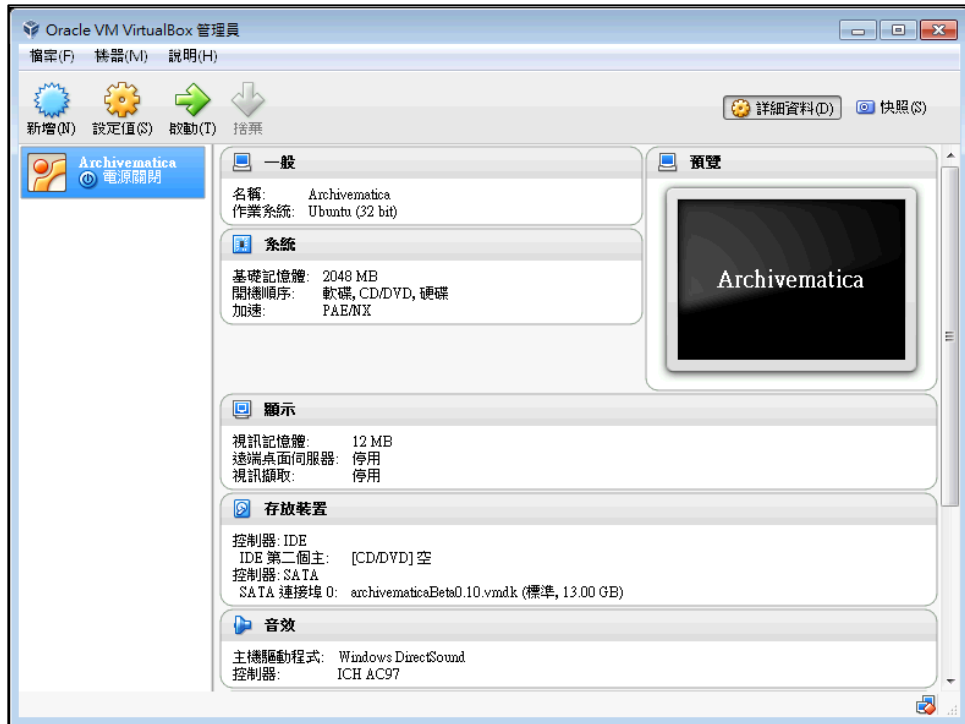
Archivematica 的基礎作業系統是採用符合自由軟體規範的 GNU/Linux 3.2.0-41，而作業系統使用的是 Ubuntu 12.04.2 LTS。Artefactual systems 提供 Archivematica 的虛擬設備(virtual appliance)²，只要將其提供的虛擬系統搭載在 Oracle VM VirtualBox 上，就有最簡單的 Archivematica 系統環境可使用，也可再依照需求設定或擴充軟體。

(一) 新增 Archivematica 虛擬機器

啟動 Oracle VM VirtualBox，先點選「新增」來建立新的虛擬機器，再輸入虛擬機器名稱並且選擇 Linux 的 Ubuntu 版本系統。記憶體大小的部分，考慮未來系統規模、硬體效能來評估與設定，建議至少配置 2048MB，以免影響虛擬機器的運作。最後，選擇現有 Archivematica 0.10-beta 版本的虛擬硬碟檔案，即可完成初步設置。







(二) 虛擬機器網路設定

為了使 Archivematica 虛擬系統的功能正常運作，主機作業系統(host OS)需要支援虛擬作業系統(guest OS)兩張網路介面卡：NAT 介面卡與 Host Only 介面卡。NAT 介面卡可讓虛擬機器不必設定 IP 就有上網功能；而 Host Only 介面卡讓 host OS 的瀏覽器得以瀏覽 Archivematica 的操作介面(dashboard)。網路設定詳細過程如下：

- (1)在 Virtualbox 選擇 Archivematica 的虛擬主機，點選「設定值」的「網路」。由於 Virtualbox 預設的網路介面卡為 NAT，因此我們只需再多新增一張介面卡，並且設定為「僅限主機」介面卡。



(2) 啟動 Archivematica 虛擬機器電源。

(3) 檢查運作中的網路介面卡，指令如下：

```
$ ifconfig
```

除了 lo 以外應有兩張名稱為「eth」+數字編號的介面卡正在運作，若沒有的話則需開啟支援的介面卡。以我們的系統為例，指令如下：

```
$ sudo ifconfig eth2 up
```

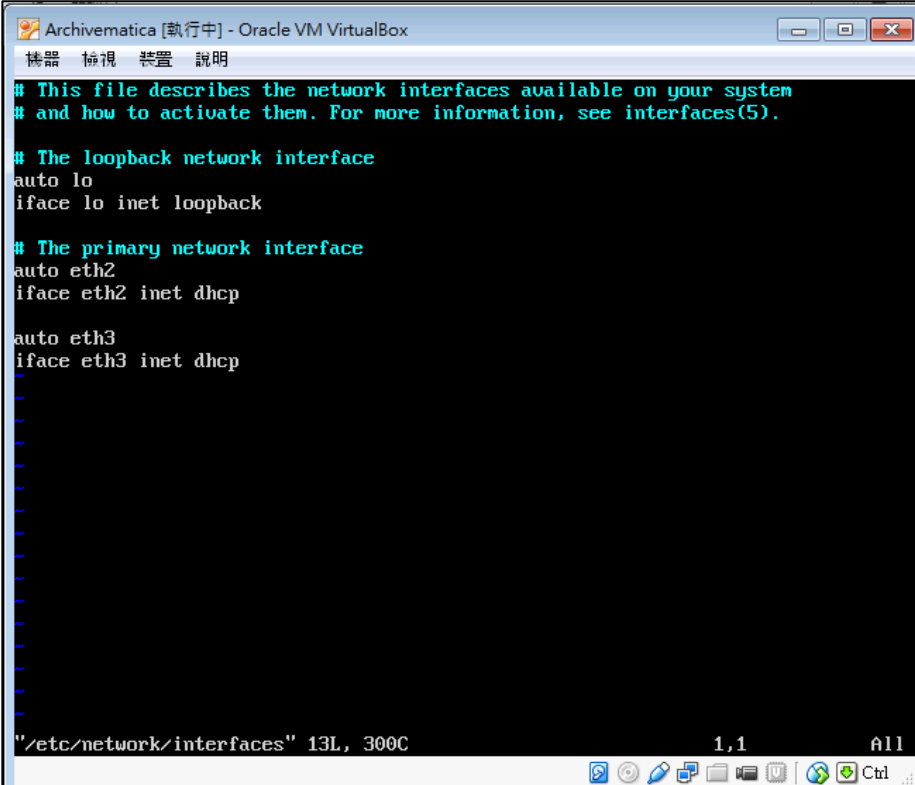
```
$ sudo ifconfig eth3 up 192.168.56.101
```

介面卡的編號會因 Virtualbox 的虛擬機器數而不同，在此例為 eth2 與 eth3 分別作為 NAT 介面卡與 Host Only 介面卡，其中在 eth3 所設定之 IP 位址(在此例為 192.168.56.101)，作為 host OS 的瀏覽器網址即可使用 Archivematica 操作介面。

(4) 編輯網路卡設定檔，編輯指令如下：

```
$ sudo vim /etc/network/interfaces
```

將檔案內的兩個 DHCP 的介面卡設定修改為剛才設定的編號。

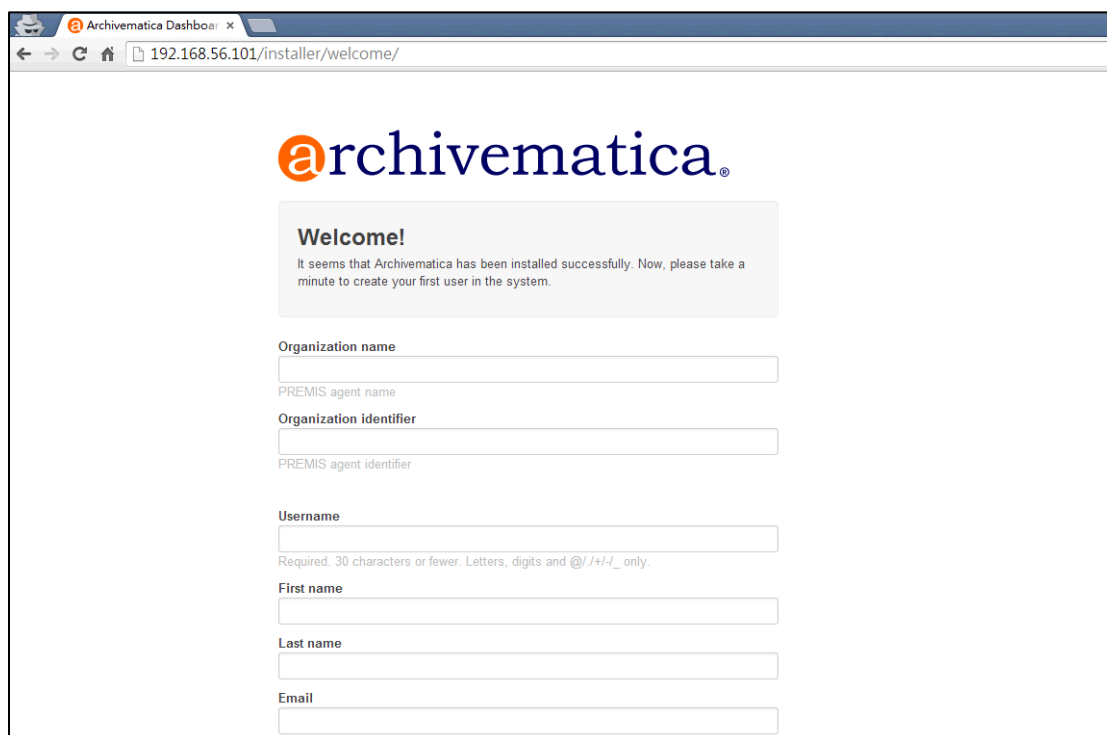


```
Archivematica [執行中] - Oracle VM VirtualBox
機器 檢視 裝置 說明
# This file describes the network interfaces available on your system
# and how to activate them. For more information, see interfaces(5).
# The loopback network interface
auto lo
iface lo inet loopback
# The primary network interface
auto eth2
iface eth2 inet dhcp
auto eth3
iface eth3 inet dhcp
...
"/etc/network/interfaces" 13L, 300C
1,1 All
```

(5) 存檔離開網路卡設定檔後，必須重啟網路才能正常運作，指令如下：

```
$ sudo /etc/init.d/networking restart
```

開啟 host OS 的任何瀏覽器連結至 eth3 設定之 IP 位址，若進入 Archivematica 的操作介面就算設定成功。

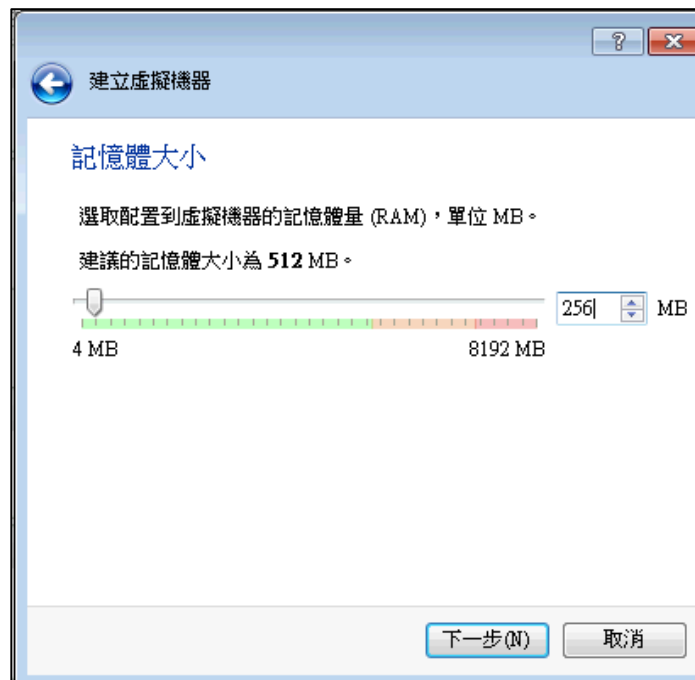
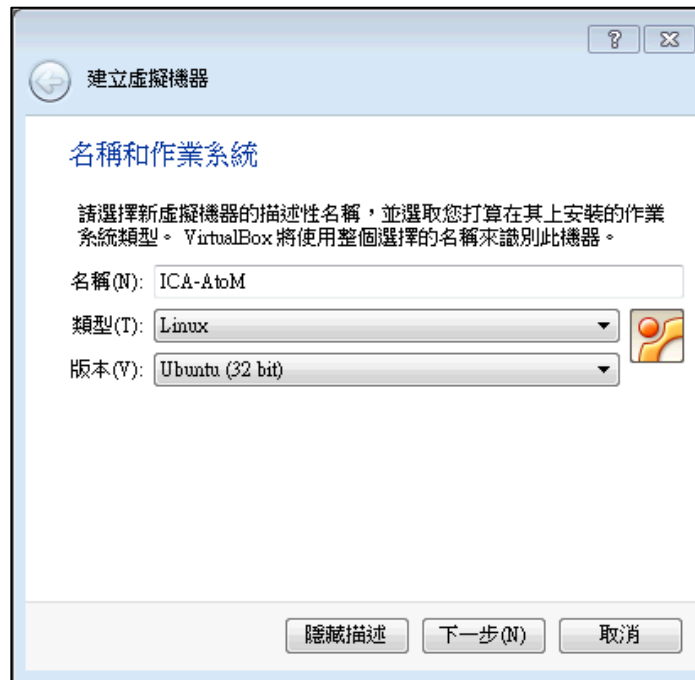


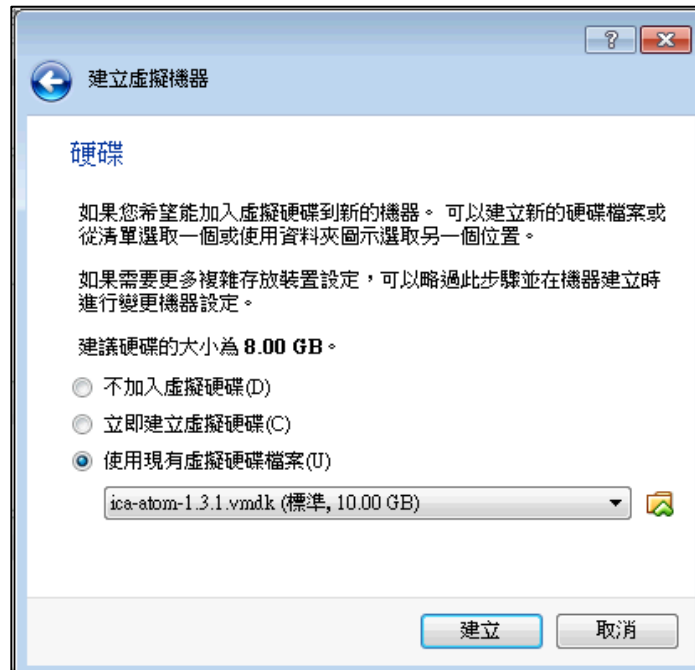
(6)最後，我們必須關機以儲存所有網路設定。建議的關機方式是透過 Virtualbox 向虛擬機器發出關機指令，步驟為：點選機器->關閉->儲存機器狀態，如此一來網路設定得以保存。

二、 建置 AtoM 伺服器

(一) 新增 ICA-AtoM 虛擬機器

ICA-AtoM 安裝方法與 Archivematica 的相似。首先在 VirtualBox 點選「新增」來建立新的虛擬機器，再輸入虛擬機器名稱並且選擇 Linux 的 Ubuntu 版本系統。記憶體大小的部分，配置基礎記憶體空間 256MB 即可。最後，選擇 Artefactual systems 提供的虛擬設備：ICA-AtoM 1.3.1 版本的虛擬硬碟檔案，即可完成初步設置。





(二) 虛擬機器網路設定

ICA-AtoM 的網路設定和 Archivematica 的大同小異，詳細步驟如下：

- (1) 在 Virtualbox 選擇 ICA-AtoM 的虛擬主機，點選「設定值」的「網路」。
設置兩張介面卡：NAT 介面卡與 Host Only 介面卡。
- (2) 啟動 ICA-AtoM 虛擬機器。
- (3) 修改網路設定檔，賦予 ICA-AtoM 伺服器 IP 位址，如此 Archivematica 才找得到傳輸目的地。我們以 eth0 和 eth1 為例，指令和修改內容如下：

```
$ sudo vim /etc/network/interfaces
```

編輯網路設定檔，修改內容如下圖所示：

```
ICA-AtoM [執行中] - Oracle VM VirtualBox
機器 檢視 裝置 說明
# This file describes the network interfaces available on your system
# and how to activate them. For more information, see interfaces(5).

# The loopback network interface
auto lo
iface lo inet loopback

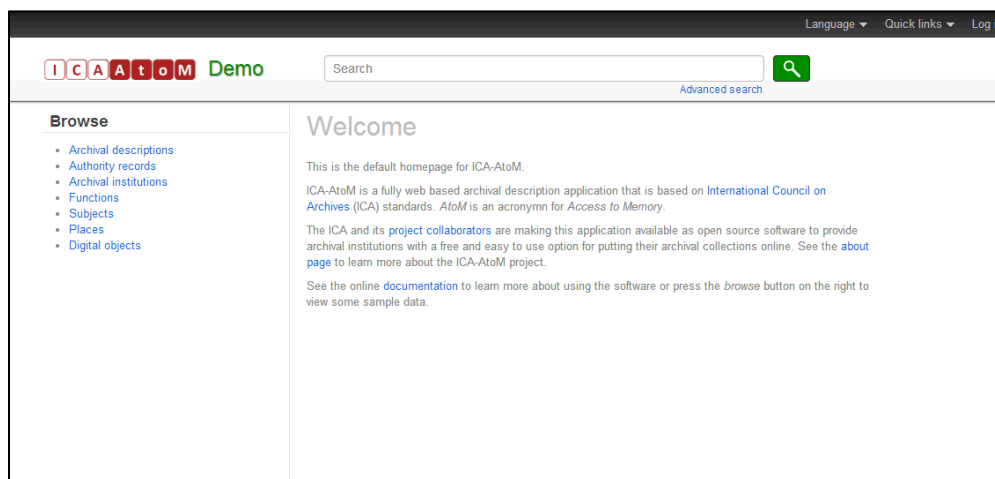
# The primary network interface
auto eth0
iface eth0 inet static
    address 192.168.56.102
    netmask 255.255.255.0
    broadcast 192.168.56.255

auto eth1
iface eth1 inet dhcp
~
~
```

存檔離開網路卡設定檔後，必須重啟網路才能正常運作，指令如下：

```
$ sudo /etc/init.d/networking restart
```


(4)開啟 host OS 的任何瀏覽器連結至 <http://192.168.56.102/>，若可進入 ICA-AtOM 的操作介面就算設定成功。



三、 建立 DIP upload 傳輸管道

上傳至 Archivematica 的資料經過處理後會產生 DIP，並且直接傳送至 AtOM 備份。Archivematica 與 AtOM 溝通橋樑是使用 Rsync 技術作加密異地備份，我們必須先讓 Archivematica 免密碼就可登入 AtOM，接著設定傳輸參數。詳細步驟如下：

(一) 伺服器連線加密設置

此步驟目的是讓 Archivematica 能自動通關到 AtOM，在這裡使用 rssh 登入的方式，比起 ssh 更加安全。

(1)我們需要產生 Archivematica 伺服器的 SSH 鑰匙，把這個鑰匙交給 AtOM 伺服器就能夠達到免密碼登入的目的。在 Archivematica 伺服器執行以下指令：

```
$ sudo -i -u archivematica
$ cd ~
$ ssh-keygen
```

產生出來的鑰匙儲存在 `/var/lib/archivematica/.ssh/id_rsa.pub`，這個檔案將會用到，因此務必備份。

(2)接著我們想要 Archivematica 以 SSH/rsync 的方式傳輸 DIP 至 AtOM，在 AtOM 伺服器安裝 rssh 並且新增一個擁有此權限的使用者：

archivematica 來加強系統安全性。在 AtOM 伺服器執行以下指令：

```
$ sudo apt-get install rssh
```

```
$ sudo useradd -d /home/archivematica -m -s /usr/bin/rsync  
archivematica
```

```
$ sudo passwd -l archivematica
```

```
$ sudo vim /etc/rsync.conf
```

進入 rsync 設定檔後，必須確保 allowrsync 那一行沒有被註解掉，修改完成後存檔離開。接下來建立一個地方儲存剛才產生的 Archivematica 伺服器鑰匙，指令如下：

```
$ sudo mkdir /home/archivematica/.ssh
```

```
$ chmod 700 /home/archivematica/.ssh/
```

```
$ sudo vim /home/archivematica/.ssh/authorized_keys
```

進入 authorized_keys 檔案編輯畫面，貼上 Archivematica 伺服器鑰匙內容，修改完成後存檔離開。最後，修改其檔案權限：

```
$ chown -R archivematica:archivematica /home/archivematica
```

(二) AtoM 參數設定

為了讓 AtoM 能夠接收 DIP，還需要設定 plugin 的參數。從 AtoM 的操作介面就可以調整參數，參數設定如下：

(1) Admin -> Plugin 的 qtSwordPlugin 設定為 enable。

(2) Admin -> Settings 的 Job scheduling 設定為 yes。

(三) Archivematica 參數設定

傳輸 DIP 必須知道 AtoM 的網址、使用者名稱與密碼，還需要設定 rsync 傳送的檔案在 AtoM 伺服器儲存的目的地，管理者可在 Archivematica 的操作介面的 Administrator -> AtoM DIP upload 裡面編輯這些參數。以我們的例子來說，參數設定如下：

```
--url="http://192.168.56.102/index.php" \  
--email="demo@example.com" \  
--password="demo" \  
--uuid="%SIPUUID%" \  
--rsync-target="archivematica@192.168.56.102:/tmp" \  
--debug
```

四、參考文獻

1. Archivematica

https://www.archivematica.org/wiki/Main_Page

2. Install-0.10-beta: Archivemata 的 virtual appliance 下載點
<https://www.archivemata.org/wiki/Install-0.10-beta>
3. ICA-Atom
<https://www.ica-atom.org/>
4. Download and run as a virtual appliance: ICA-Atom 的 virtual appliance 下載點
https://www.ica-atom.org/doc/ICA-Atom_virtual_appliance
5. Archivemata 0.10 的 Administrator manual
https://www.archivemata.org/wiki/Administrator_manual_0.10

出國報告（出國類別：國際會議）

參加 2014 年國際圖書館協會聯盟 (IFLA)年會與參訪報告

服務機關：國立政治大學圖書資訊與檔案學研究所
姓名職稱：薛理桂教授、陳慧婷計畫助理
派赴國家：法國
出國期間：2014 年 8 月 15 日至 8 月 26 日
報告日期：2014 年 9 月 1 日

出國成果報告書

出國人員	薛理桂 陳慧嫻	出國日期	103 年 8 月 15 日至 103 年 8 月 26 日，共 11 日
出國地點	法國	出國經費	科技部計畫、計畫結餘款
報告內容摘要			
<p>2014 國際圖書館學會聯盟(The International Federation of Library Associations and Institutions 年會於法國里昂召開，本年度的主題為「圖書館、公民和社會：為知識而彙聚」，其核心理念為圖書館使公民和社會更加強大。 IFLA 每年年會都吸引世界各國圖書館人員及資訊人員參與，以增進與各國圖書館間的交流合作，了解全球圖書館及資訊服務最新發展趨勢。本次活動參加 IFLA 會議並展示國科會計畫之海報，並配合中華民國圖書館學會安排之參訪行程。參訪當地國家圖書館與國家檔案館，藉以了解法國圖書館與檔案館的實際作業服務、空間運用及館藏內涵等等，將有助於日後教學。</p>			
<p>2014 國際圖書館學會聯盟(The International Federation of Library Associations and Institutions 年會於法國里昂召開，本年度的主題為「圖書館、公民和社會：為知識而彙聚」，其核心理念為圖書館使公民和社會更加強大。因會議參與人數眾多，並同時段有多場次同時進行，以下將略述數個場次的重要內容。</p> <p>一、1001 個生前必看的圖書館</p> <p>IFLA 公共圖書館委員會在 2014 年年會上，宣布 1001 個生前必看的圖書館計畫，會議中介紹了傳統圖書館以典藏和閱讀為主，然而現代圖書館則是與民眾生活融為一體，除典藏、閱讀之外，也具有教育和休閒娛樂的功能，並提供芬蘭、瑞士、西班牙、澳洲、英國等數個入選的圖書館照片，目前僅 100 多個圖書館入選，歡迎世界各國圖書館提出申請。相關資料詳網站：http://1001libraries.wordpress.com/</p> <p>二、特藏的特別空間—圖書館建築及典藏善本書與手稿的設備</p> <p>第一篇論文關於位在法國史特拉斯堡的國家與大學圖書館，是法國第二大的歷史圖書館，典藏超過 300 萬件文件，包含古代、中世紀及現代館藏。館舍建於 1895 年，成為城市的地標建築，在 2010 年時閉館進行館舍大整修，此計畫稱為“BNU Nouvelle”，目的不只保存獨特的歷史館藏，也希望以新的方式呈現給大眾。因此館舍計畫改為開放式博物館，允許團體參訪以發掘豐富與多元的圖書館館藏，包含紙草紙、勳章、西方與東方的手稿、地圖和繪畫、雕像、傢俱以及珍貴的印刷版本等等。該館在歷經 4 年修復後，將於 2014 年 11 月重新開放。第二篇為美國 Salisbury University 的 Nabb Research Center for Delmarva History and Culture，討論關於特藏手稿在館員小心維護館藏，與提供公眾使用之間如何取得平衡。作者提到並沒有可統一使用於各個圖書館的方法，因每個圖書館都有其特定任務與價值。第三篇論文關於美國西密西根大學於 2013 年 10 月啟用的 Zhang Legacy Collections Center，該中心典藏 3 萬立方公尺的地方歷史圖書、手稿與檔案資料。西密西根大學為美國密西根州第四大公立大學，並為研究型大學，此中心共有 6 名全職館員及 12 名學生助理與志工，文章論述在嚴格的溫濕度控制下，</p>			

如何規畫建立管理此中心的工作區與典藏區。第四篇論文關於法國 Interuniversity Library of Sorbonne 論述如何管理 1897 年至 2016 年的館藏空間規劃。

三、電子書的數位保存

第一篇論文是關於法國國家圖書館延長電子書寄存的作法，電子書的標準格式含 PDF 或 EPUB，以及內嵌後設資料(ONIX)，在法國有越來越多民眾使用電子書，因此如何長期保存電子書成為一大挑戰。主要作法參考數位文件以及傳統寄存方法，包含傳統紙本書的電子化以及原生電子書，預計產生一套標準工作流程，於 2015 年開始實施。第二篇論文為德國國家圖書館為快速發展的電子書建立標準流程，希望使出版商能以自動化介面，負責傳遞線上資源。並以標準化的後設資料與介面為基準，使數位物件的完整性與品質可驗證，以提供大眾使用，並可轉入圖書館的數位保存系統。第三篇為日本 National Diet Library (簡稱 NDL)，在著作權法限制下，如何收集線上出版品。NDL 有義務收集特定檔案格式(包含 PDF, EPUB, DIASY)、具書目編碼 (ISBN, ISSN, DOI)。但實際上 NDL 只能收集免費使用的以及未受到 DRM(Digital Rights Management)保護的電子出版品。該文探討如何解決此難題，圖書館需要與主流出版社簽約(例如 3 年)，以短暫寄存線上出版品。最後一篇論文為中國國家圖書館如何保存電子書。

四、知識、資訊與民眾：圖書館的社會價值

關於巴西聖保羅州政府與聖保羅大學發展名為 ACESSA SP 計畫，探討老人參與此計畫的情況，比較 2008、2009 與 2012 年所收集的統計資料，結果指出老年人在行為、使用和數位素養各方面都有增加的趨勢，參與 Facebook、web2.0、部落格、YouTube、Twitter 等，可了知巴西老年使用網路的情況與習慣。

五、IFLA 中的原住民議題：加強圖書館與資訊中的文化回應與權責性

第一篇論文關於紐西蘭原住民教育，以四種主要原則(反思、學習與教學、熱情與精神、關係)，並以學術和支援品質服務作為加強。以學術服務來支援被正規教育所忽略的原住民教育。第二篇為台灣中研院報告數位典藏計畫，將中研院民族所典藏品如何數位化、如何設計與發展異質性後設資料，以及如何在跨平台上做共通檢索，以提供建置全球性原住民數位圖書館作為參考。第三篇為美國自 1970 年代起和原住民進行口述歷史之經驗，並探討如何保存與行銷這批口述歷史資料。第四篇為加拿大 University of Manitoba，圖書館館員分享如何協助同學學習與了解原住民文化的經驗，以加強原住民社群和學術圖書館之間的連結。最後一篇為美國華盛頓大學如何透過內容管理系統(CMS)，解決個別原住民以及原住民社群的需求。

六、PERSIST 計畫—數位保存與選擇

第一篇論文探討數位文化遺產的數位環境議題，由於數位內容的產生、消費與取得都有高度改變的情況，對於數位經濟有極大的影響，從產生到保存應以怎樣的方式進行有效管理。第二篇論文為探討南非與非洲的數位遺產內容保管者和產生者，在處理政治、社會、理想與技術上的掙扎，關於哪些資料要長期保存，與資訊控制、權力與資源都有相關的議題。最後一篇為法國國家圖書館如何保存原生數位資料。

七、海報參展

本次參與 IFLA 會議，參展海報主題為數位長期保存系統建置：以 Archivematica 為例，在每日海報展示說明期間，參觀海報展的人數非常眾多，同時也可發現各國各層級圖書館在製作海報上，其創造力與巧思非常驚人與值得學習。我們在海報展示時，也遇到英國、新加坡、中國等各國圖書館員對我們的海報感到興趣，並進行討論，是非常難得的經驗。

八、參訪心得

在此次參與 IFLA 大會時，參與中華民國圖書館學會參訪行程，參觀里昂市立圖書館，該圖書館位於里昂火車站對面，圖書館於 1972 年完工，圖書館大門出口即是百貨公司，另一邊出口面對火車站，交通非常便利，因應逛街的民眾需求，圖書館在週日也開館服務。該圖書館的特色為收集里昂百年來地方出版品，並擁有法國最豐富的中文圖書，其館藏以 1921 年中國政府在里昂設立「里昂中法大學」之圖書與檔案為基礎，並受里昂大學委託寄存中法大學 473 位學生之學籍檔案。里昂市立圖書館目前與國內的臺北市立圖書館為姐妹館，國內臺北市圖的館員曾到該館進行三個月的交流。該館與國內的國家圖書館、台北市立圖書館、上海、廣州圖書館進行圖書交換，並接受法國漢學家的私人捐贈，每一位捐贈者每一次捐贈以一個全宗作為處理（此種作法與檔案管理方式相同），圖書與檔案編號僅做流水號的編號排架，在編目系統中會詳細編目與分類。中文館藏豐富，但僅有兩位工作人員，同時精通中文與法文，將中文圖書以法文編目，中文圖書的保存與流通也是他們的工作，每年平均有 30 位民眾到館使用，圖書流通量一年約有 1500 冊，一位民眾平均使用 50 冊圖書。

除了參觀圖書館之外，本人與研究生也參觀里昂市立檔案館，由於該館未對外開放，無法一窺檔案館內部館舍，但該館建於地鐵站旁，交通非常便利，在保留舊有檔案館建築之下，賦予新的外觀設計，並將百年舊大門鑲嵌於牆上，非常特別的將舊有建築和新建築融為一體。另外，在里昂還有羅納省檔案館，完工於 2013 年，以三大棟的金色建築所構成，設計理念體現出檔案館保存內容的價值，該館將於 2014 年 9 月 15 日開館。

從參訪此三館中，我們發現法國圖書館或檔案館皆建置於交通便利之處，且將民眾閱覽區和密集書庫分開建置，通常有一棟矮建築（約 5 層）作為民眾閱覽中心，一棟高建築（約 17 層）作為密集書庫，此種作法使館員在管理時也較為便利，閱覽區僅有較常使用的館藏，而年代久遠或較為珍貴的書籍則放至於密集書庫，以得到良好的保存環境與管理。

九. 建議

經由此次參與國際會議與參訪圖書館與檔案館後，提出三項建議；

一. 國內應興建國家級檔案館

目前國內尚未有國家層級之檔案館，此次參訪法國的國家檔案館舊館與新館，規模宏偉，可知該國對於檔案資產之重視，可供國內參考，應儘速規劃與興建國家層級的檔案館。

二. 國內應興建國家圖書館新館

在法國塞納河畔興建有四座如開啟的書本的國家圖書館，可知該國對於典藏該國文化資產的國家圖書館之重視。國內目前雖有位於中山南路的國家圖書館，但已面臨館藏飽合之困境，建議效法法國之舉措，儘速規劃與興建國家圖書館新館。

三. 國內應興建地方層級檔案館

法國各地區都興建有地方的檔案館，此點在國內還付諸闕如，未能典藏各縣市所產生的地方檔案，建議各地方政府應儘速規劃與興建地方層級的檔案館。

Long-Term Preservation Strategy and System Implementation for Digital Archives

Li-Kuei Hsueh , Hui-Pin Chen

Graduate Institute of Library , Information and Archival Studies
National Chengchi Univeristy, Taipei, Taiwan

Abstracts

The Taiwan e-Learning and Digital Archives Program has been implemented for a decade led to the creation of many digital archives. Since both information technology and storage media have developed so quickly, the major issue for these archival repositories is how to preserve the digital results for the long term through migration methods. This study uses literature reviews to realize the current status of digital long-term preservation strategy and software. It also interviewed with archival staff of the Institute of Taiwan History, Academia Sinica to find their need and as an example. This study uses Archivemata as digital preservation software in order to establish a digital preservation strategy and determine a better solution for Taiwan's archival repositories.

Keywords: Long-term Preservation; Digital Archives

Introduction

- Purpose: Realize digital archives, how to preserve the digital results over the long-term
- Case Study : Archives of the Institute of Taiwan History, Academia Sinica
- Literature review : realize the current status of digital long-term preservation strategies and software
- Interview with archival staff of the Institute of Taiwan History, Academia Sinica to find their needs and as an example.
- Preservation software: Archivemata

Archivemata

- Archivemata is a free and open-source digital preservation system managed by Artefactual Systems.
- The Archivemata project is collaborates with the UNESCO Memory of the World's Subcommittee on Technology and with several archives, museums and university libraries.
- Archivemata uses a micro-services design pattern that allows an archives institution to process digital objects in compliance with the OAIS functional model.
- Advantages:
 - Archivemata suit for archivists and librarians who have minor technology ability, it can lower the barriers to best-practice digital preservation.
 - Maximize long-term investments for the benefit of colleagues, users and professional community as a whole.

Interview with the archivists of the Institute of Taiwan History, Academia Sinica

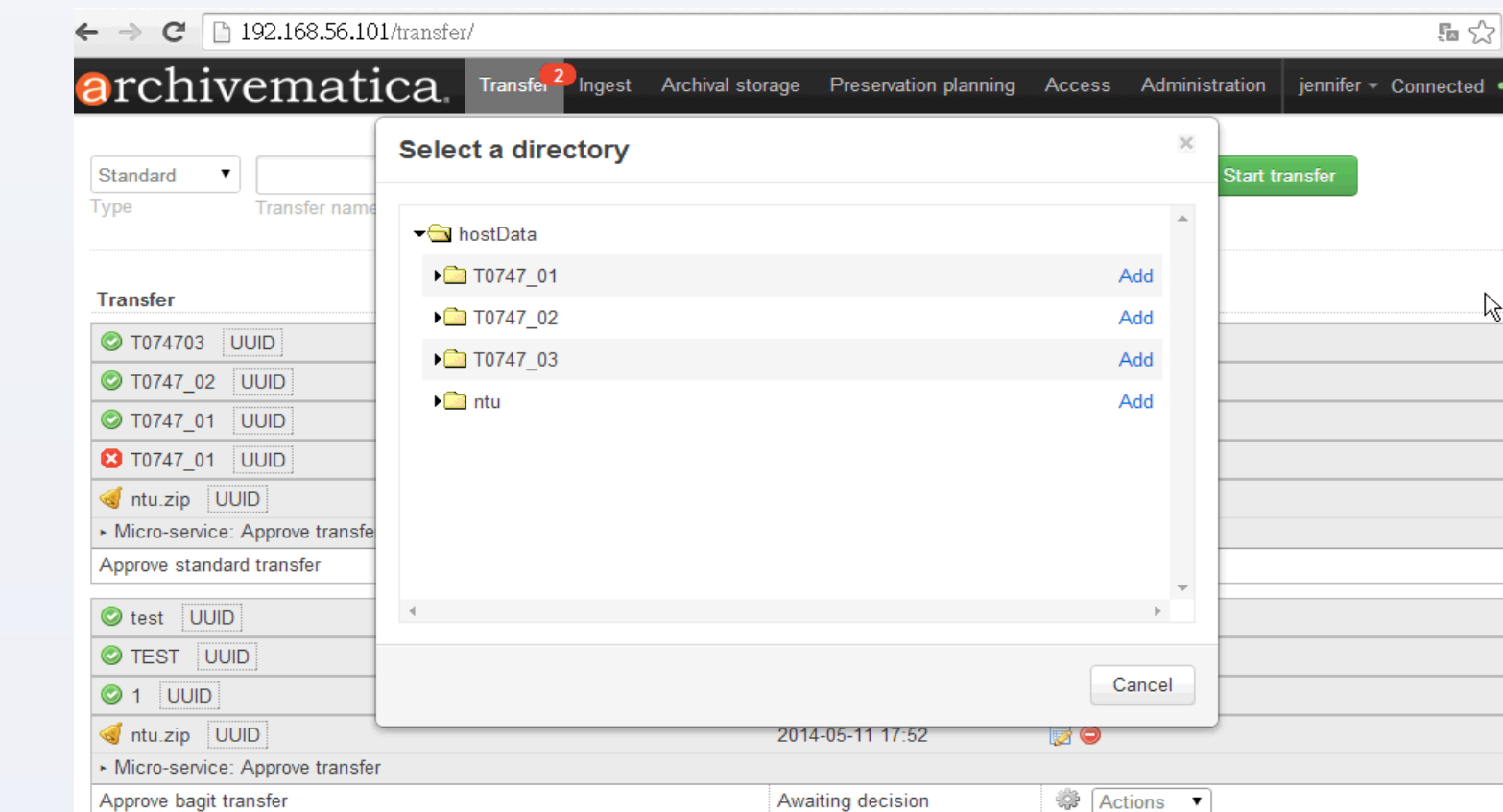
- Because of limited space, archives of the Institute of Taiwan History, Academia Sinica collaborate with other institutions, so 90% of the collection is digital archives.
- The archives have about 10 million digital files.
- Digital format : TIFF format for archives, JPEG 300 dpi for books, 600dpi for paintings. JPEG 300 dpi will be uploaded to a digital repository, and then have the form transferred for browsing and thumbnails.
- Storage media use DVD, hard disc drives for TIFF and JPEG file. Magnetic tape and RAID for JPEG files.
- Offsite backup at the Academia Sinica Computing Centre.
- Do not record preservation metadata, only list storage date and object identifier number. Digital object and descriptive metadata stored separately.
- The archives do not have a preservation policy or preservation strategy.

Case Study

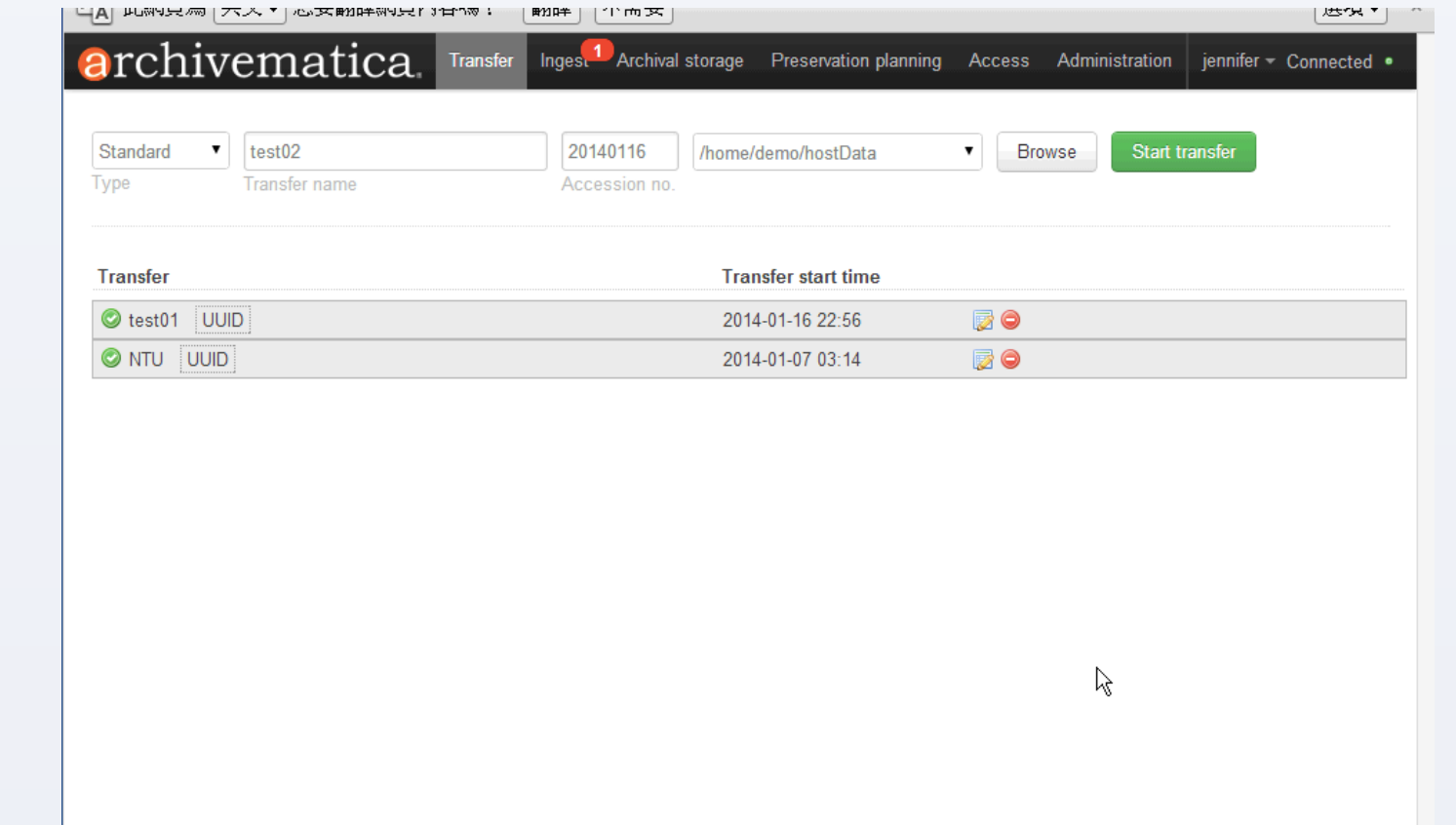
- Fonds: Yoshioka Kisaburo
 - Control level: use series level:Diaries of Yoshioka Kisaburo(T0747_01) as first example.
 - Control level: use series level: Public records of Yoshioka Kisaburo (T0747_02) as second example.
 - Control level: use series level: Photos of Yoshioka Kisaburo (T0747_03) as third example.
- Use Archivemata as digital preservation software, pretest from SIP to AIP.
- Package for Ingest
 - One package for one series, containing digital images(jpeg) and metadata CVS file
 - Export DC-Metadata XML from the repository
 - Produce Metadata CSV files manually for Archivemata
- Upload to Archivemata
 - Open Oracle VM VirtualBOX
 - Put ingest packages to Archivemata folder
 - Open Archivemata software
 - Select upload folders
 - Select ingest type: standard/ unzipped Bag/Zipped Bag/ Dspace/Maildir
 - Entry transfer name and Accession no.

Upload to Archivemata

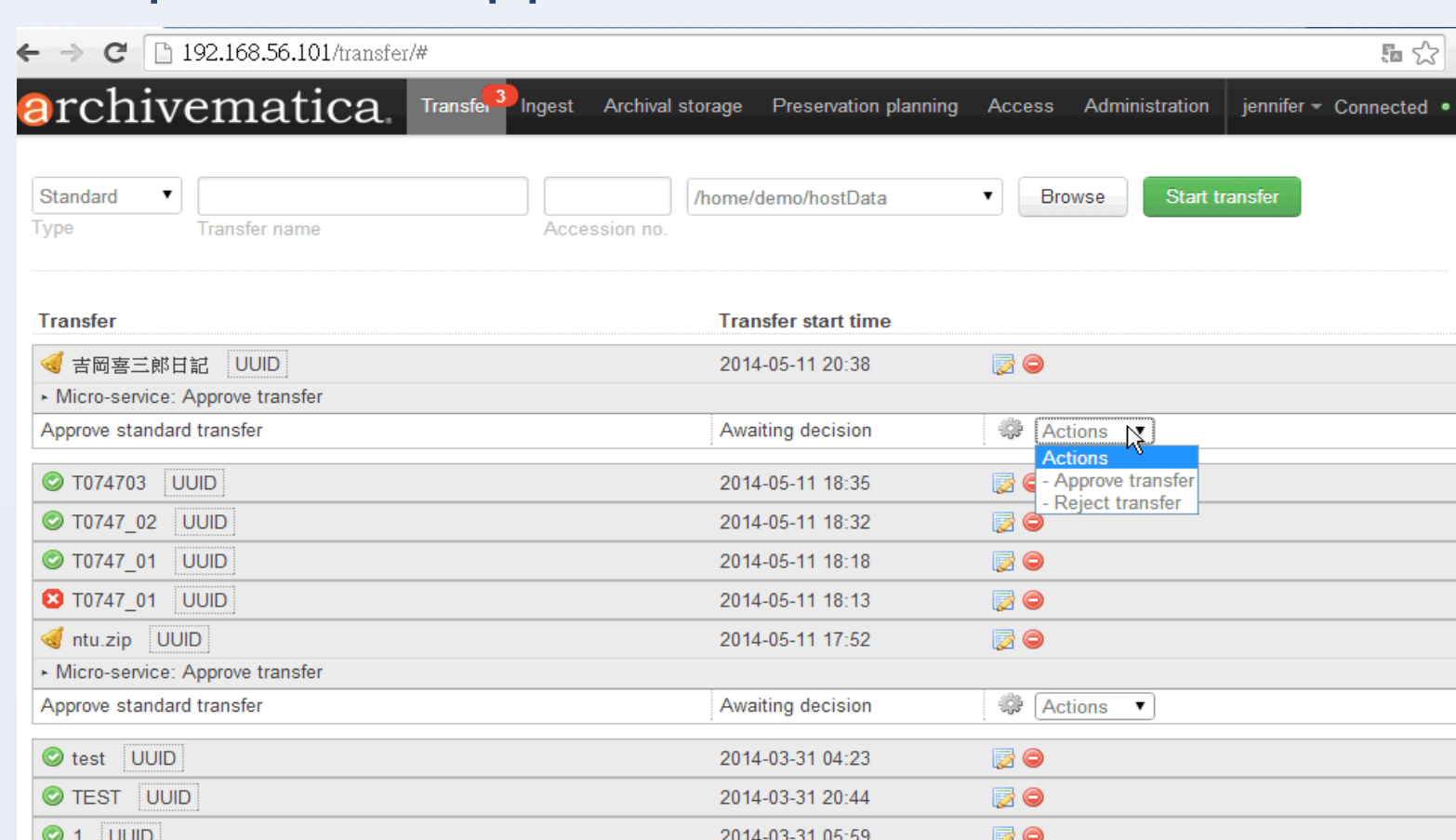
Step 1 Select upload folders/



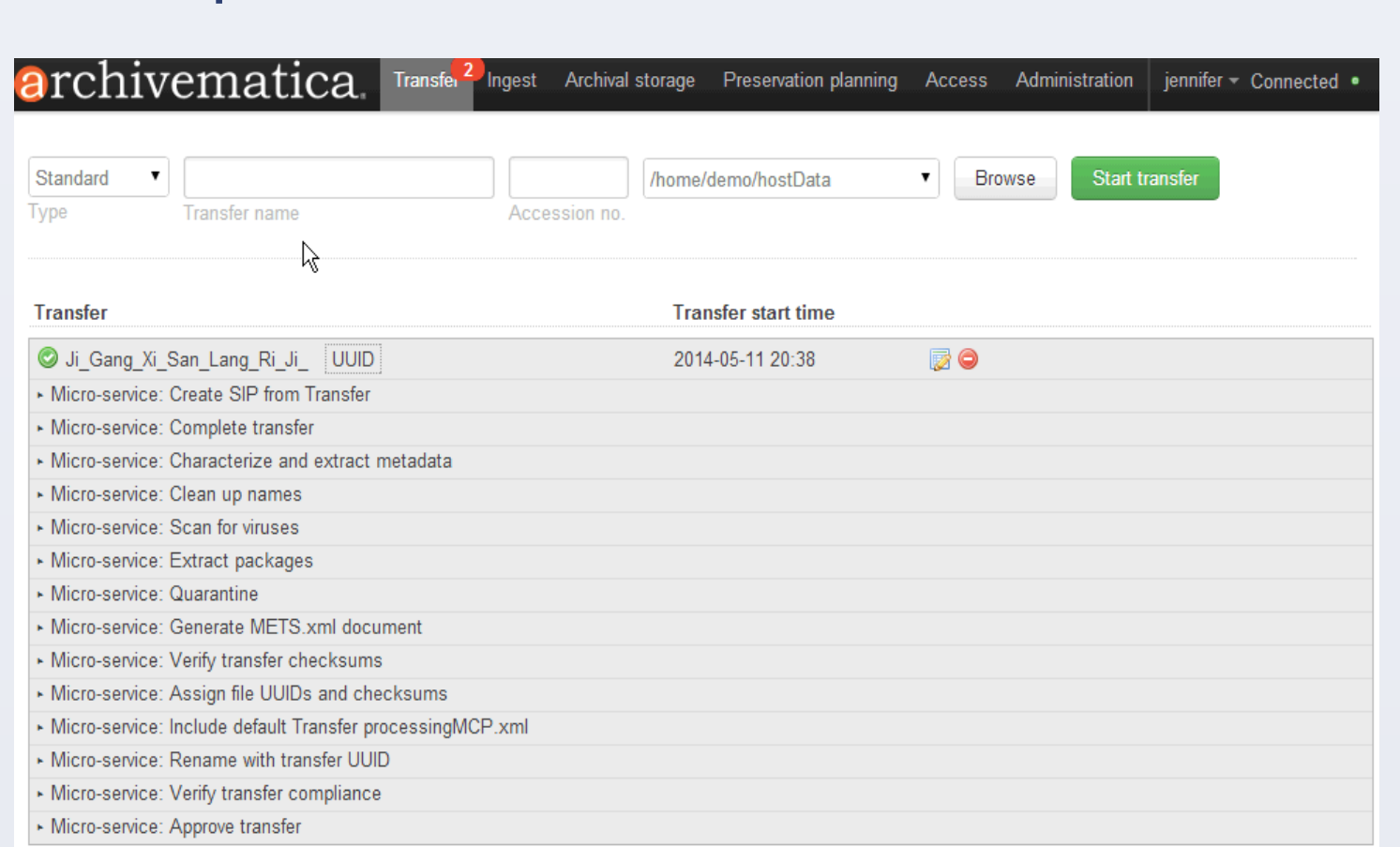
Step2 Select Standard Type/ Entry transfer name



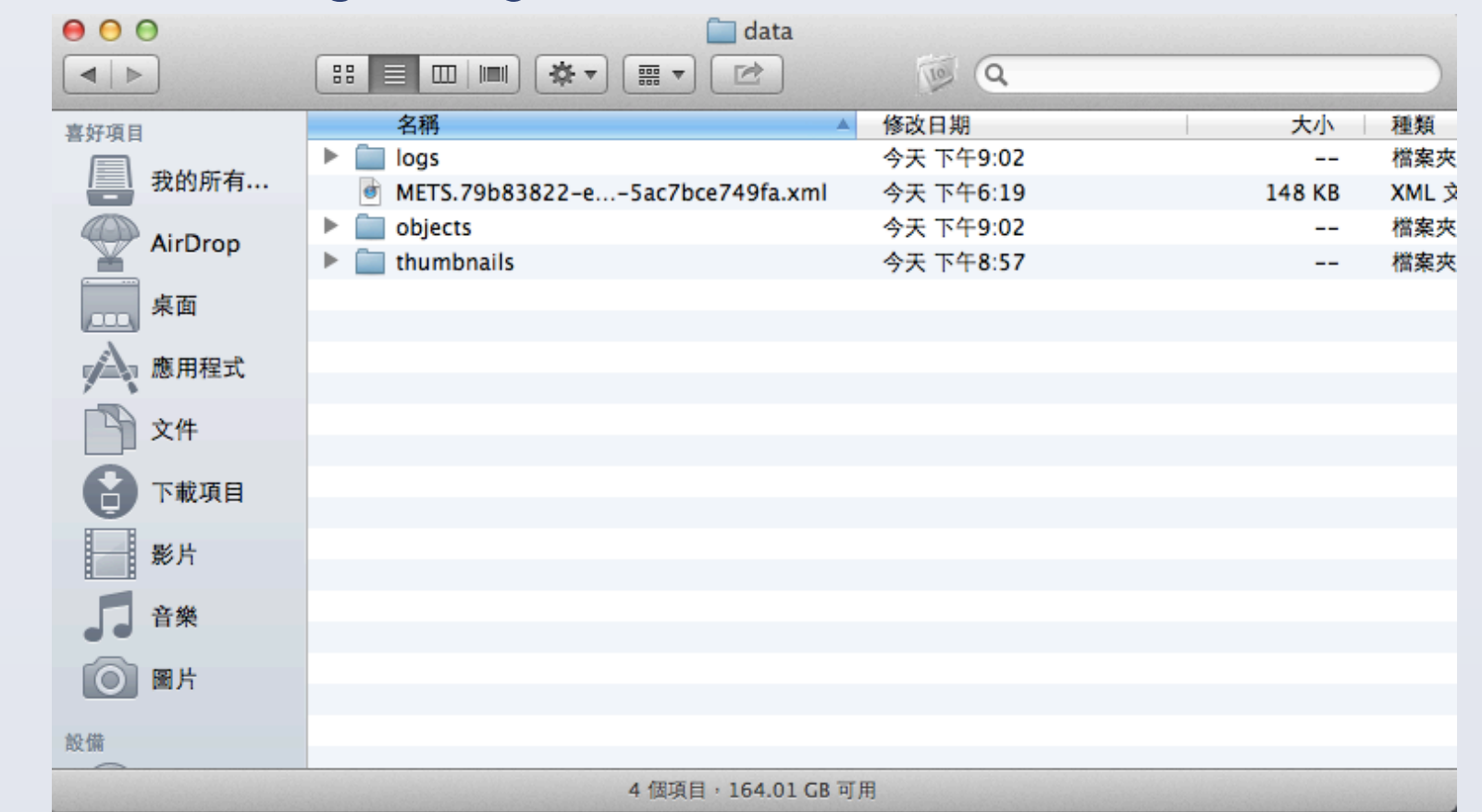
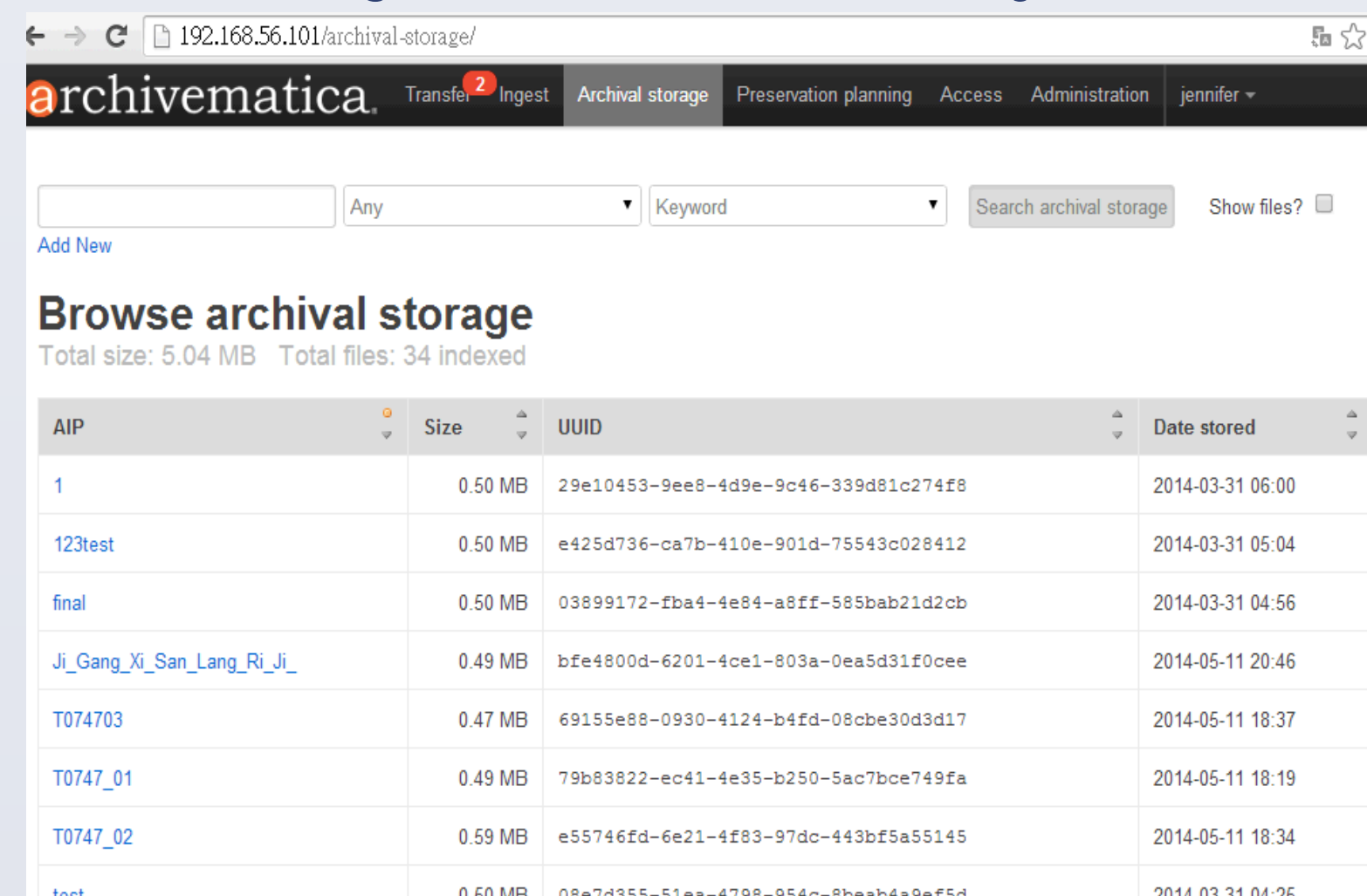
Step 3 Click Approve transfer/



Step 4 automated micro-service for SIP



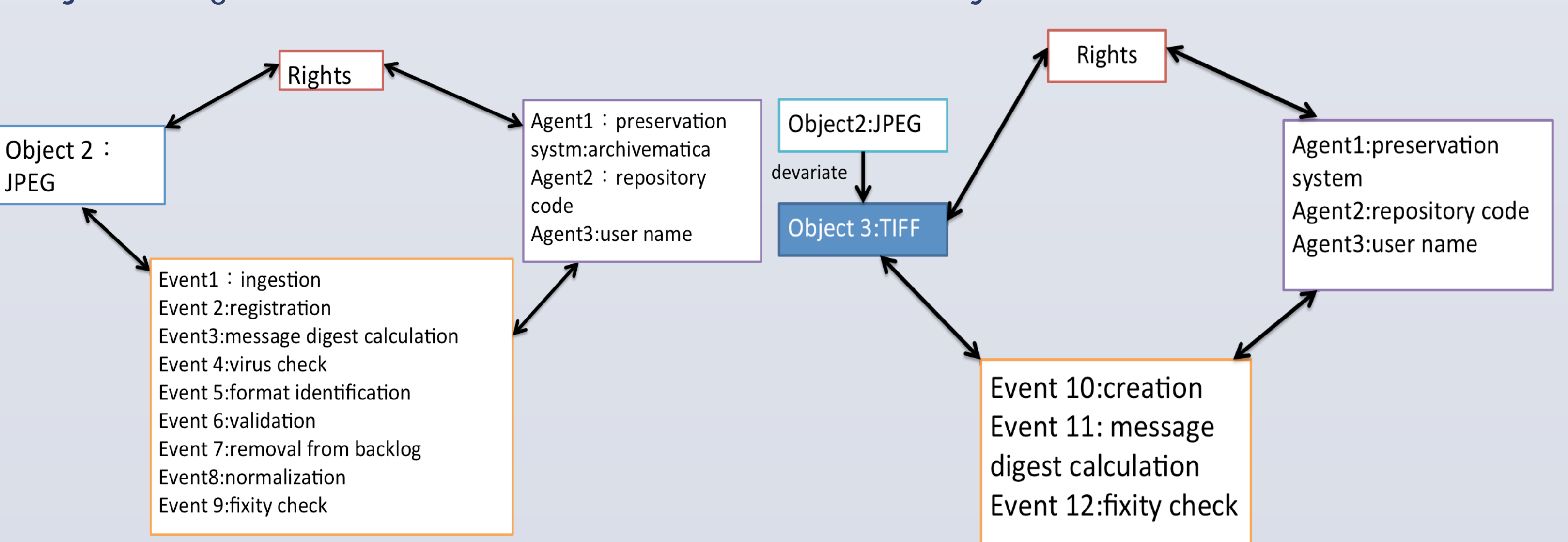
AIP Package :METS file directory, and three folders : logs, objects, and thumbnails



PREMIS Metadata Data Model

Object: Original JPEG

Object: Normalization TIFF



Conclusion

- Archives should adopt Archivemata as digital preservation software, even just adopt default functions, archivists can complete digital preservation management tasks successfully.
- Archivemata can a designate unique identifier for each object, event, and agent.
- Archivemata stores digital objects and descriptive metadata together, solving the problem of preserving digital objects and metadata separately.
- Archivemata uses METS and PREMIS Standard record digital preservation procedures, tools and dates, so archivists can manage digital preservation more easily.
- If an archives adopts Archivemata as its digital preservation software, the archiving should first develop a digital preservation strategy.
- Depending on the archives policy and context, an archives can enter Rights Metadata.
- Archives need IT professional to set up Archivemata, find solutions and display Chinese characters.

Reference

- Archivemata(2014). *User manual 1.1*. Retrieved from https://www.archivemata.org/wiki/User_manual_1.1
- Garderen, P.V., Jordan, P., Hotten, T., Mumma, C., McLellan E. (none) . The Archivemata Project. Retrieved from http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC_Van_Garderen_et_al_26_Workshop1.pdf
- The Archives of Institute of Taiwan History, Academia Sinica.(none) Yoshioka Kisaburo fonds. Retrieved from http://tais.ith.sinica.edu.tw/sinicafrsFront/search/search_detail.jsp?xmlid=0000251776

Acknowledgements

The work described in this paper was sponsored by the Ministry of Science and Technology under MOST Grants: MOST 102-2410-H-004-220-

科技部補助計畫衍生研發成果推廣資料表

日期:2014/11/20

科技部補助計畫	計畫名稱: 數位典藏長期保存策略及系統建置之研究: 以中研院台史所檔案館為例
	計畫主持人: 薛理桂
	計畫編號: 102-2410-H-004-220- 學門領域: 圖書文獻學
無研發成果推廣資料	

102 年度專題研究計畫研究成果彙整表

計畫主持人：薛理桂		計畫編號：102-2410-H-004-220-					
計畫名稱：數位典藏長期保存策略及系統建置之研究：以中研院台史所檔案館為例							
成果項目		量化			單位	備註（質化說明： 如數個計畫共同 成果、成果列為該 期刊之封面故 事...等）	
		實際已達成 數（被接受 或已發表）	預期總達成 數(含實際已 達成數)	本計畫實 際貢獻百 分比			
國內	論文著作	期刊論文	0	0	0%	篇	
		研究報告/技術報告	0	0	0%		
		研討會論文	0	0	0%		
		專書	0	0	0%		
	專利	申請中件數	0	0	0%	件	
		已獲得件數	0	0	0%		
	技術移轉	件數	0	0	0%	件	
		權利金	0	0	0%	千元	
	參與計畫人力 (本國籍)	碩士生	2	0	0%	人次	
		博士生	1	0	0%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
	研討會論文	2	0	100%			
<p>1.Hsueh, L.-K.*, & Chen, H.P. (2014).The Long-term Preservation Strategy and the System Implementation of Digital Archives.In:World Library and Information Congress : 80th IFLA General Conference and Assembly(IFLA2014). Aug.16-22, 2014. Lyon, France.</p> <p>2. 陳慧嫻、薛理桂 (2014)。數位長期保存政策擬定之探析。2014 年第十二屆海峽兩岸圖書資訊學學術研討會，天津南開大學主辦。</p>							

	專利	專書	0	0	100%	章/本	
		申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力 (外國籍)	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		

其他成果 (無法以量化表達之 成果如辦理學術活 動、獲得獎項、重要 國際合作、研究成果 國際影響力及其他協 助產業技術發展之具 體效益事項等，請以 文字敘述填列。)	無						
--	---	--	--	--	--	--	--

	成果項目	量化	名稱或內容性質簡述
科 教 處 計 畫 加 填 項 目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與(閱聽)人數	0	

科技部補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

達成目標

未達成目標（請說明，以 100 字為限）

實驗失敗

因故實驗中斷

其他原因

說明：

原先在計畫書中欲以 DAITSS 系統作為實驗測試，但因 DAITSS 系統程式在安裝時發生嚴重問題，無法完成安裝。本計畫後來找到新的長期保存系統 Archivematica，同為免費可測試軟體，使得本研究最終仍可順利完成。

2. 研究成果在學術期刊發表或申請專利等情形：

論文：已發表 未發表之文稿 撰寫中 無

專利：已獲得 申請中 無

技轉：已技轉 洽談中 無

其他：（以 100 字為限）

1.Hsueh, L.-K.*, & Chen, H.P. (2014). The Long-term Preservation Strategy and the System Implementation of Digital Archives. In: World Library and Information Congress : 80th IFLA General Conference and Assembly (IFLA2014). Aug.16-22, 2014. Lyon, France.

2. 陳慧嫻、薛理桂（2014）。數位長期保存政策擬定之探析。2014 年第十二屆海峽兩岸圖書資訊學學術研討會，天津南開大學主辦。

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

由於國內執行數位典藏計畫多年，以產出多元豐富的成果，如何能將這些珍貴的數位典藏成果長期保存，是一大重要議題。本研究除以焦點座談了解國內重要典藏單位進行數位長期保存的現況與困境之外，並以文獻探討了解國外長期保存政策與策略制定的要項與內容。此外，本研究以 UNESCO 和歐美檔案館、博物館與圖書館共同合作開發之 Archivematica 長期保存工具軟體，作為測試，此軟體不僅為 open source 軟體，且以簡易的圖形化介面，使一般不閤電腦程式的館員，也能輕易的藉此軟體管理數位典藏的成果，並可產出符合國際保存性後設資料標準的 XML，並可檢索和存取長期保存的成果，如此可有效解決僅以異地備份作為保存策略的缺陷。唯一的缺憾是，目前 Archivematica 軟體並沒有中文化，在處理中文字體上也需後續專業技術人力

的努力，若能改善此缺點，對各館以此軟體進行數位長期保存，必定很有助益。