

漢代簡牘數位典藏系統之設計與實作

黃國倫 李家豪 何建明

中央研究院資訊科學研究所

{ kulun, chahao, hoho }@iis.sinica.edu.tw

摘要

數位資訊化與全球資訊網正快速改變傳統研究方式與知識獲取來源。數位典藏就是希望將原本難以廣泛分享的知識來源，經由整理數位化後，提供文化、教育、社會、產業等各種發展應用。

本研究以數位典藏計畫為基礎，從人文及技術面的分析與設計方法，試圖提出一適合漢代簡牘數位典藏之系統架構，並在實作中整合多媒體、缺字、全文檢索等相關技術。在文中描述整個數位化過程與資訊系統開發經驗，提供各數位典藏單位與相關研究人員參考，期望能將專業領域之內隱知識 (Tacit) 外顯化 (Explicit)，讓知識分享、再利用及加值更加容易。

Keywords

數位典藏、漢代簡牘

1. 前言

竹簡木牘是中國古代紙張發明以前最主要的書寫材料，百餘年來，中國西北地區漢晉烽燧遺址或各地戰國秦漢墓葬出土大量的簡牘，生動地記錄了西漢中晚期到東漢初期，當地軍民在軍事、法律、教育、經濟、信仰以及日常生活各方面活動的情形，舉凡漢代的吏制、養老、撫恤、秩奉、農墾屯田、背算、上計、邊塞防禦等制度或算數、曆法等問題，均可從所

藏簡牘得到深具價值的研究素材，是國內一批價值非凡的文化資產[1]。

「數位典藏國家型科技計畫」在民國 91 年 1 月 1 日正式成立，承襲行政院國家科學委員會「數位博物館計畫」、「國家典藏數位化計畫」、「國際數位圖書館合作計畫」三個計畫的經驗，依據國家整體發展，重新規劃而成。計畫的首要目標是將國家重要的文物典藏數位化，建立國家數位典藏；進而以國家數位典藏促進我國人文與社會、產業與經濟的發展。

漢代簡牘數位典藏系統係屬於「數位典藏國家型科技計畫」機構計畫「中央研究院歷史語言研究所文物數位典藏計畫—拓片與古文書分項」的典藏品資料庫之一，自九〇年代初期起由中央研究院歷史語言研究所簡牘整理小組，以紅外線攝影儀、掃描器、電腦，重新釋讀所藏簡牘，並進行影像、釋文資料數位化。因此，進而規劃建置資料庫檢索與管理系統。

建置工作首先由簡牘整理小組負責漢代簡牘後設資料的分析，並在中央研究院後設資料工作組 (Metadata Architecture and Application Team, MAAT)[2] 支援下，完成後設資料的規劃、分析與實施，並協助標準化、結構化工作流程與內涵分析。最後交由中央研究院資訊科學研究所數位典藏技術發展與支援組依照簡牘整理小組所提出之系統功能與介面設計，並參

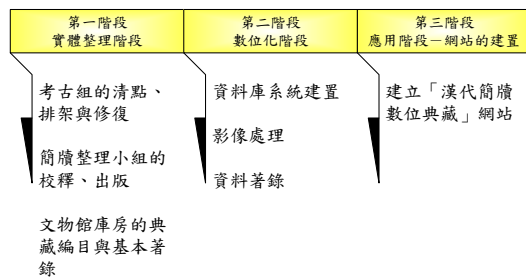
考後設資料工作組製作之「漢簡後設資料需求規格書」[3]，再整合國內外相關資訊技術而開發之數位典藏系統。

透過本系統的運作，除了提供漢簡基本資料、釋文以及彩色照、黑白照、紅外線照、遺址平面圖等數位影像之網路線上瀏覽，也整合典藏資料、人員資料、相關研究性資料及內部作業流程，以提供完善的漢代簡牘數位典藏環境，由此增進漢簡研究的便利性，更進一步提升資料整合性與整體服務的品質。

2. 背景與文獻探討

2.1 漢代簡牘數位典藏工作流程

中研院歷史語言研究所漢代簡牘數位典藏工作流程[1]，分為三個階段執行，如圖表 1：



圖表 1：漢代簡牘數位典藏工作流程

1. 第一階段—實體整理階段：

以 1930 年瑞典考古學家貝格曼 (Folke Bergman) 在甘肅、內蒙古境內額濟納河流域發掘的居延漢簡為主，對於實體整理主要在漢簡的清點及排架工作，以製成文物典藏清冊，其中也進行部分殘簡的修復工作。

史語所文物館成立於 1986 年，除了商周青銅器、甲骨、陶器、玉石器、骨器等考古文物之外，漢代簡牘也是

該館的重要典藏品之一，為便於文物管理、展示、修復、科學鑑定、調件等工作，近年也著手進行漢簡典藏的編目及基本資料的建檔。

自 1988 年起，漢簡整理小組 (之後改稱「簡牘整理小組」) 開始著手進行所藏簡牘的校釋工作，以出版圖錄與原簡並針對所藏簡牘的校釋工作。除了以資料卡記錄新校釋文與文字清晰評比的資訊外，並進行釋文的電腦打字與造字。

2. 第二階段—數位化階段：

早在數位典藏計畫之前簡牘整理小組已進行漢簡影像數位化或資料庫建置，自 1992 年便採購紅外線影像處理器及電腦影像強化及儲存系統，以紅外線攝影儀所見字跡，經影像強化處理後，儲存於光碟作為資料之保存。接著也陸續建置，「簡帛金石資料庫」與「居延漢簡補編圖象檢索系統」，在此基礎之下，進而規劃漢代簡牘數位典藏系統。

影像數位化也是這階段的重點，除了原有照片檔案的整理工作外，還有照片、底片掃描，並以原簡影像攝影，而產出有紅外線數位影像與彩色數位影像。這些圖檔也會分別儲存於中央研究院計算中心之巨量儲存系統以及歷史語言研究所數位典藏計畫之磁碟陣列、並加以浮水印處理。

3. 第三階段—應用階段：

在這階段主要在建立「漢代簡牘數位典藏」網站，內容包括重要文物的展示，讓一般民眾更了解漢代簡牘，並且整合所有漢簡相關資料庫系統，提供以檢索、瀏覽之功能。最後

提供歷史年表、地圖工具和相關文獻資料，延伸漢代簡牘的研究空間。

2.2 漢代簡牘 Metadata 分析

在不同的領域所建立的數位資料型式常常有相當大的差異，為了考量不同的使用者需求，以及資料庫存取資料的效率等因素，再加上不同廠商提供的資料庫功能與格式上的差異，資料庫的標準化是一個相當困難的問題[4]。因此後設資料制定成為一項重要的工作。

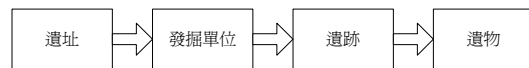
漢代簡牘後設資料的規劃、分析與設計工作是由簡牘整理小組與後設資料工作組負責，主要利用描述、結構與管理三大類型，以及語義、語法與詞彙三大屬性以描述數位典藏品的內涵與特徵。並建立後設資料分析與著錄規範，以提供國內外學界漢代簡牘資料庫開發之參考範例，更能與國內外相關研究單位、博物館資料庫或線上展覽資料之互通與交換。整理上述內容製作成「漢簡後設資料需求規格書」。

2.2.1 考古後設資料關係

中研院歷史語言研究所漢代簡牘大部份典藏於中研院歷史文物館之中，館中除了商周青銅器、甲骨、陶器、玉石器、骨器等考古文物之外，漢代簡牘也是該館的重要典藏品之一。因此在漢代簡牘後設資料設計上，一方面依照考古知識架構中將考古調查與發掘所發現之遺物、所有過程之記錄與影像檔案，以數位化的方式有效的保存於電腦系統中，使得考古調查與發掘之現場與過程，能儘可能完整的記錄與保存，進而提供考古

研究之的參考。另一方面，分析簡牘完整之特性，保留資料需求欄位，以呈現出典藏特色與價值。

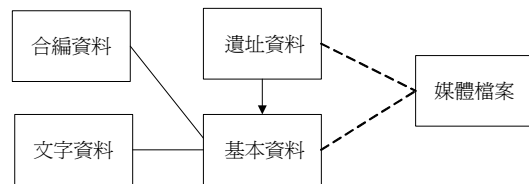
考古後設資料是在「考古發掘遺物、照片、記錄與檔案數位典藏計畫」中所制定，將其區可分為四個層次，分別為遺址、發掘單位、遺跡與遺物四層次[5]，其關係如下：



圖表 2：考古資料架構圖

2.2.2 漢代簡牘後設資料與內涵關係

漢代簡牘後設資料因考量與考古研究的相關性，故設計包含於考古知識架構中，這樣作法雖然增加後設資料分析的複雜度，但卻可以保留彼此領域的特性以整合成更完整的知識架構。而漢代簡牘僅有考古中的遺址與遺物層次資料，並且在遺物層次之中加入了簡號、釋文、說明校記等文字資料，以完整描述簡牘的內涵與特徵。故其內涵關係如下：



圖表 3：漢代簡牘內涵關係圖

2.2.3 CDWA

Metadata Standard 的對照是作為典藏品 Metadata 欄位和國際 Metadata Standard 的接軌，以及與國際間其他主題計畫作 XML 資料交換的關鍵機制。漢代簡牘 Metadata 是選擇與 CDWA[6]對照。CDWA 是由 AITF(the Art Information Task Force)所發展的後

設資料，其主要的描述目標是藝術品、建築、物件群組以及視覺化或文本的代理物(surrogates)。主要類目總共有 26 個。其設計理念主要是從藝術史研究的角度出發，其欄位希望可以滿足藝術史研究的需求，提供給研究者一致、可信的內容。另一方面，CDWA 所定義的架構也提供共享和交換的功能，使各地的典藏機構可以交換共享典藏品資訊[7]。

2.3 簡帛金石資料庫[8]

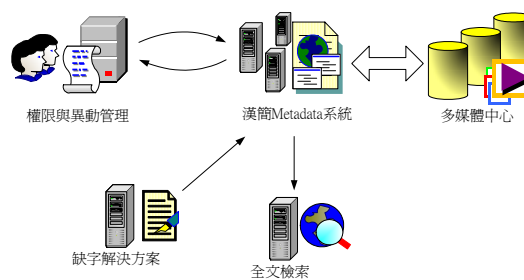
由中央研究院歷史語言研究所簡牘工作小組所建立之全文、書目資料線上檢索系統，其內容包括史語所藏「居延漢簡」釋文資料與其他地區出土簡牘、帛書、銅鏡銘、碑刻等釋文資料及簡帛研究相關論著目錄，共六十多種，提供資料檢索與研究上的便利。



圖表 4：簡帛金石資料庫

3. 系統架構

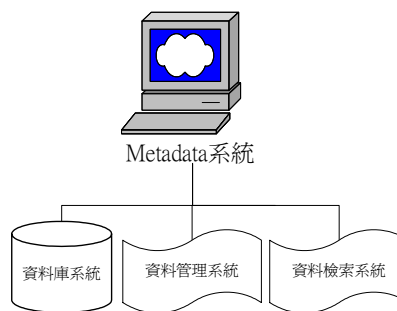
漢代簡牘數位典藏系統的設計理念在於提供數位典藏管理與服務環境所需的整合功能，讓各種工作流程，皆可透過此系統完成。其架構如下：



圖表 5：系統架構圖

3.1 漢簡 Metadata 系統

Metadata 系統是漢簡數位典藏系統的重要核心，所有後設資料的管理維護工作皆在 Metadata 系統中完成。設計上依照「漢代簡牘後設資料需求規格書」的分析結果並配合簡牘整理小組的工作流程，包括文物典藏之庫房管理（涉及文物之科學鑒定、保存狀況、修復維護、展覽出版、調件等機制），提供表單著錄與資料檢索等功能。尤其在系統資料庫實作上充分配合後設資料需求規格書中系統化、架構化與標準化的資料格式，並作為網頁撰寫時參考使用，其內容如下：



圖表 6：Metadata 系統

3.2 多媒體中心

一般數位典藏系統，不僅是文字內容，往往圖片、聲音、影像也有高度的保存需求。因此數位典藏系統對多媒體檔案管理的工作更顯重要。故須建立一套多媒體檔案管理系統負責

檔案儲存與整理工作。

在漢代簡牘數位典藏系統中，結合中研院資訊所多媒體中心[9]，可一貫化完成影像上載與轉檔工作，而網頁呈現上，以圖文共同展示設計，如圖表 7，並可依不同放大倍率瀏覽單一圖像。

若直接連線多媒體中心則可使用進階功能，包括檔案分類管理、線上轉檔或批次上載等。



圖表 7：圖文共同展示設計

3.3 權限與異動管理

就功能與資料而言，「漢代簡牘數位典藏系統」須符合不同類型使用者之操作需求。因此系統應針對使用對象和處理權責分別提供存取權限之設定，故在使用者管理上可分為二類：

1. 工作人員管理：工作人員主要負責系統的相關管理工作，其中的權限控制項目包括了系統管理功能、著錄表單功能、著錄代碼維護功能、著錄作業管理功能、資料檢索功能。
2. 資料庫使用者管理：通常是開放給一般使用者提供線上資料瀏覽檢索功能。

異動管理主要是提供典藏系統稽核的機制。紀錄使用者何時操作典藏

系統及針對那一筆典藏資料所做的異動。不但可以作為資料追蹤上重要的證據，若以使用者或日期為檢索條件，則查詢出來的記錄也可以成為績效評估的統計報表。

3.4 缺字解決方案

由於簡牘上面書寫有許多文字，內容記錄了西漢中晚期到東漢初期，當地軍民之軍事、法律、教育、經濟、信仰以及日常生活的情形。這些文字往往不是目前常用的字體，故無法於電腦系統中輸入、顯示與查詢這些缺字。

目前一般對於處理古籍數位化都有獨自建立的造字檔以解決缺字問題，但使用相當不便利。因此漢代簡牘數位典藏系統提出的解決方案則採用中研院資訊所開發之漢字構形造字機制，將資料庫中之缺字部分以構字式儲存，其中包括了輸入、顯示、查詢等的處理工作。

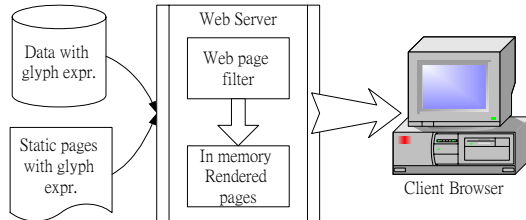
1. 缺字輸入：

解決方法有二，一是安裝中研院資訊所文獻處理實驗室之漢字構形資料庫[10]；二則經由網路缺字查詢工具[11]，複製所需缺字至漢簡系統。兩者差異在於漢字構形資料庫必須事先安裝於 USER 端，可與使用者之輸入法、文字編輯器相互結合，有較友善之輸入介面；經由網路查詢缺字，使用者端雖不需安裝任何軟體，卻會受限於網路之不確定因素，但兩種輸入方式卻不相抵觸，可自行依使用者電腦環境選擇使用。

2. 缺字顯示：

採用 Server-Side 及時過濾轉換

(On-the-fly filtering)，將含有缺字之動靜態網頁，在 Response 給使用者端之前，會先經過 Web-page filter 將缺字轉換成圖形。而處理過的網頁會暫儲於 Server 端之記憶體中，最後再傳送給使用者端，如下圖：



圖表 8：網路缺字顯示系統架構

3. 缺字查詢：

只要將使用者欲檢索之缺字部分以著錄相同方式輸入，當在查詢時僅會以字碼比對搜尋，故可得到包含缺字內容之結果。

3.5 全文檢索

漢簡後設資料中保留了釋文、說明校記等文字資料，因此利用一套獨立的全文檢索系統，檢索該全文欄位。

檢索系統主要依據中研院資訊所 KP (Knowledge Portal) 之全文檢索系統規劃而成，在此檢索系統中，須先依使用者需求定義出欲檢索之漢簡 Metadata 欄位，並以 XML 方式描述 [12]。

考量了典藏系統與檢索系統間之資料同步化問題，其處理流程如下：

1. 當典藏系統資料有異動 (Insert、Update、Delete) 時，系統會向全文檢索系統發出同步化要求，並送出異動資料的 XML。
2. 檢索系統收到同步化要求後，將收到之 XML 格式的資料解析後，紀錄其對應到的欄位資訊。

3. 以 Index Engine 進行分析與建立索引並紀錄至資料庫。

如此，使用者即可透過全文檢索查詢漢簡釋文或說明校記之內容。

4. 系統實作

4.1 Metadata 著錄

Metadata 著錄是數位典藏重要工作之一，將整理過的資料及研究成果依照 Metadata 著錄規範記錄在典藏系統中。以 Metadata 元素並依照簡牘整理小組建議，將系統著錄功能區分成「基本資料」、「合編資料」、「文字資料」、「媒體檔案」、「遺址資料」、「文獻資料」六類表單。

漢代簡牘是由「基本資料」、「合編資料」、「文字資料」組成。「基本資料」表單為漢簡品名、年代、遺址資訊、形製、材質、度量等基本資訊。「合編資料」表單為簡冊復原與簡牘綴合的版本記錄。「文字資料」表單為漢簡不同版本之釋文以及相關著錄文獻之資料。

「遺址資料」表單主要著錄出土簡牘遺址之相關資訊，當漢簡基本資料中若有指定其所屬遺址，則可透過遺址記錄識別碼連結方式帶出遺址相關資料。

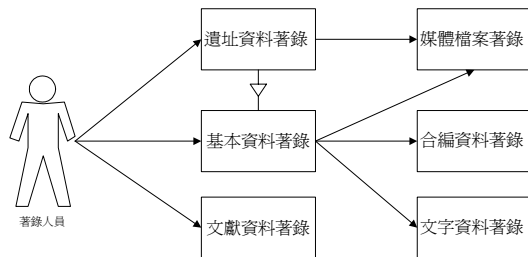
「媒體檔案」表單的設計是為了滿足漢簡及遺址中對於多媒體檔案典藏管理功能，在表單中包括影像檔名、業務種類、影像類型、典藏位置等影像資訊之記錄。

「文獻資料」是獨立於漢簡 Metadata 之外的表單，由於上述五個表單中相關著錄的完整出版資料需求，而記錄包括文獻之簡稱、作者、

篇名、書名、期刊名、出版資料之內容，以提供各表單連結。

4.1.1 表單著錄

漢代簡牘以「基本資料」表單為主體，而其他表單著錄時，系統須檢查該筆記錄的基本資料是否已被建立，以確保資料的完整性，故在著錄流程設計上，著錄人員須優先著錄「遺址資料」、「基本資料」及「文獻資料」。當著錄「基本資料」時，由於表單中會帶出所屬遺址的資料，故須在之前完成遺址資料著錄才能順利連結，見圖表 9。



圖表 9：表單著錄流程

4.1.2 媒體檔案著錄

由於漢代簡牘數位典藏系統中，媒體檔案僅包括文字資料與影像資料。雖然影像品質包括典藏級、參考級及網路瀏覽級，而檔案格式也包括 TIFF 及 JPG，但透過與多媒體中心的整合後，可以容納上述類型檔案。媒體檔案著錄工作，當新增一筆媒體檔案時，須指定該記錄的資料種類，可以設定為遺物(漢簡)或遺址，並且透過查詢方式選擇資料編號以減少人工輸入上的錯誤，如圖表 10。



圖表 10：查詢方式選擇資料編號

由於媒體檔案包括文字資料與影像資料，透過網頁可以著錄文字內容。然而影像實體檔案則要採取不同著錄方式，考量到網際網路的負荷，因此建議若原始圖檔小於 1MB，可直接透過網頁上傳檔案。若檔案過大或數量過多時，則可利用批次轉檔方式，交由系統管理者直接轉入。

4.2 系統功能

除了著錄之外，漢代簡牘數位典藏系統也實作了許多功能，包括檢索功能、文字直式排版、校對報表、使用權限、異動管理及代碼管理。

4.2.1 檢索功能

主要提供漢簡基本資料、釋文以及彩色照、黑白照、紅外線照、遺址平面圖等數位影像之網路線上查詢瀏覽，並依照不同使用者特性，提供不同的檢索功能(簡易檢索、進階檢索)。

4.2.2 文字直式排版

由於簡牘是以前的書寫材料，因此上面的文字也提供了重要的研究史料。在「文字資料」表單中包括了不同版本之釋文資料，透過網頁格式呈現上，須依照原簡牘的排列格式呈現，並能支援文字直式排版，如下：



圖表 11：文字直式排版

4.2.3 校對報表

透過檢索方式，將查詢結果以報表格式呈現，並提供分頁的報表列印功能，供著錄者校對資料或製作書面資料之使用。

4.2.4 使用權限

該功能提供使用者帳號管理、使用者群組管理及功能模組管理，如圖表 12。故可針對使用對象和處理權責分別提供存取權限之設定，以此決定系統建檔人員與系統管理人員之操作功能，以有效防止非合法授權者侵入系統內取用資料。而在使用者管理可依每個人員設定不同的工作權限，並且可將相同層級之人員設定成群組帳號，而透過群組權限設定即可改變該群組內所有人員權限設定。



圖表 12：使用權限管理

4.2.5 異動管理

異動管理為內部工作紀錄之統計資料，該功能結合使用權限，並保留了資料異動之操作與更新的人員與時間，故可透過不同的查詢條件製作成報表，以提供績效考核之依據。其欄位資訊如下：

欄位名稱	欄位說明
ID	自動編號
記錄識別碼	異動的資料編號
記錄類型	該資料所屬表單
日期	異動發生之時間
執行政序	異動所執行的操作
使用者	執行異動的人員
群組名稱	該人員的群組
備註	其他相關記錄資訊

4.2.6 代碼管理

依後設資料需求規格書所提供的代碼表，以各表單著錄所使用的代碼分開個別維護。可以新增、修改、刪除每個代碼的項目，並能依照考古與漢簡之需求提供不同的代碼內容。



圖表 13：

5. 結論

數位典藏系統建置是一連串複雜工作組成，幸好本計畫中有簡牘整理小組－負責系統規畫、後設資料工作組－協助後設資料分析及標準化，以及技術發展與支援組－系統開發及技術整合，所以漢代簡牘數位典藏系統雛型才得以完成。雖然過程中遇到不少問題，尤其是漢簡須整合至考古資料架構中這增加系統實作上的困難度，但也借此機會能進一步了解，做為下一階段考古數位典藏系統建置工作的事前準備。

本研究中將整個系統建置的規劃與設計，包括作業流程、後設資料分析、系統架構、系統實作等，提供各數位典藏單位與研究人員在數位典藏工作執行上的參考。

未來的研究上，可以從幾個部份去加強：

- 雖然在系統中已將漢簡整合至考古後設資料架構之中，故能達到資料互通的優點。但未來若能將範圍擴大，也整合至其他相同領域，或是不同領域，那將對典藏品知識的呈現上會更加完整。這也是數位典藏聯合目錄建置中重要的工作。
- 系統安全性機制一直是重要指標，雖然在本系統已實作了權限與異動的管理功能，但未來系統運作後對於數位資料的保護需要更加嚴謹，一方面配合各單位人員審核的作業流程；另一方面提供各系統間彼此不同的權限設定。若從資料的角度來看，那影

像的保護也更加重要，故浮水印機制也會一研究重點。

參考文獻

- [1] 陳秀慧, “中研院歷史語言研究所漢代簡牘數位典藏工作流程簡介”,
<http://content.ndap.org.tw/result/process/09theme-as01/hanjan.htm>
- [2] MAAT, 數位典藏後設資料工作組
<http://www.sinica.edu.tw/~metadata/>
- [3] MAAT, “拓片與古文書數位典藏計畫漢代簡牘後設資料需求規格書 version 0.2”
- [4] 何建明、黃世昆、莊庭瑞、李德財, “典藏數位化資訊環境之探討”, Oct 2000
- [5] 林玉雲, “保存消失中的資訊——一個適合考古遺物的典藏系統架構”, 第一屆數位典藏技術研討會, July. 2002
- [6] CDWA, Categories for the Description of Works of Art,
<http://www.getty.edu/research/institute/standards/cdwa/>
- [7] MAAT, 後設資料參考規範,
<http://www.sinica.edu.tw/~metadata/standard/standard-frame.html>
- [8] 簡帛金石資料庫,
<http://saturn.ihp.sinica.edu.tw/~wenwu/search.htm>
- [9] 中研院資訊所多媒體中心,
<http://ndmmc.iis.sinica.edu.tw/Index.jsp>
- [10] 漢字構形資料庫,
<http://ckip.iis.sinica.edu.tw/CKIP/tool/>
- [11] 網路缺字輸入法,
<http://140.109.18.63/word/s.html>
- [12] 郭沛顯、賴震宇、林宣華、何建明, “檢索系統和數位典藏系統間的通訊協定”, 第一屆數位典藏技術研討會, July. 2002