

第三章 數位典藏系統建置

黃國倫、林金龍

第一節 系統建置評估

第二節 系統分析

第三節 系統設計

第四節 資料庫開發

第五節 使用權限管理

第六節 漢字編碼處理

傳統的典藏管理工作需要依賴大量的人力，藉由導入資訊系統，可以有效降低管理成本並提高工作效率。相同地，數位典藏在保存重要數位化資料的過程中，也將藉由各項資訊技術的協助，解決典藏工作上所遭遇的問題。尤其對建置數位典藏系統而言，不但涉及資訊系統的開發，亦牽涉到各典藏內容的專業領域、文化背景、法律授權與國際標準等面向，更必須從各層面分析數位典藏的需求、應具備的功能及可能面臨的瓶頸等問題。因此本章將介紹從系統建置過程中，如何透過詳實的分析與設計，以及典藏相關技術的整合，以建置一個完整的數位典藏系統。另外，亦會深入系統核心，針對資料庫開發、使用權限管理與漢字編碼處理等重要議題加以說明，以確保資訊系統建置完成後，在未來可以保持運用彈性和長久服務。

第一節 系統建置評估

一、數位典藏環境分析

目前國內從事數位典藏系統或技術開發的單位不在少數，尤其是在知識經濟概念的熱潮帶動下，不論政府單位或是私人企業，都相繼投入此領域。然而此領域所需整合的知識、技術項目相當多，因此在數位典藏環境中存在著一些仍待克服的問題。以下舉例說明之（范紀文、何建明、李德財，民90）：

1. 內容專家與技術人才之合作

數位典藏工作需要結合典藏知識的內容專家，與資訊科技的技術人才。但由於彼此間專業領域不同，平時交流、互動機會亦少，以致於在目標的認定上、語

言詞彙的表達上，差異極大。在這種彼此缺乏了解的情況下，如何讓技術人員正確的分析、規劃、設計出符合內容專家所需要的典藏系統，是數位典藏系統開發的一大問題。加上數位典藏環境中使用到多項資訊技術，如網路工具、資料庫管理、影像視訊處理、資訊安全及資料搜尋等技術，即使是資深的技術人員，也往往無法全數熟悉及掌握。現實環境中，大部份的技術雖已發展成熟，也有實際的產品，但這些各自獨立的技術核心，需與現有的系統整合，以發揮最佳的效果。這些整合的技巧需要厚實的技術訓練，與系統開發整合的實務經驗相輔相成，並非一蹴可及。因此在發展數位典藏時，需要有擅長溝通之內容專家，及有系統整合經驗之技術人才共同參與，才能發揮事半功倍的效果。

2. 後設資料分析與系統分析之結合

由於後設資料分析的目的在于將複雜的資料內容予以組織成有系統的資料結構，並從內涵意義中建立資料結構間的關係。因此在後設資料規格中，會針對典藏單位的著錄需求，規劃出適合資料欄位與屬性，以符合系統管理與檢索上的應用。所以在進行系統分析前，必須先著手後設資料分析，以掌握後設資料的著錄需求後，接下來再根據著錄需求與功能需求出發，進行整體典藏系統的規格分析。為了避免後設資料分析的結果與系統設計的內容產生矛盾，必須在系統分析時，參照後設資料功能需求書，以充分了解後設資料的需求，促使後設資料分析、系統開發兩方面達成共識、建立溝通的橋樑。

3. 典藏系統的通用性

為了達成不同單位間典藏資料能夠交換，因此有許多組織制訂後設資料標準。雖然目前幾個常用的後設資料國際標準皆已發展成熟，但若實際要將不同類型的典藏資料，納入單一的後設資料標準架構中，卻是一個極富挑戰性與爭議性的問題。尤其各領域表達知識的架構不一，所使用的後設資料標準與系統功能需求亦有所不同。因此，在系統設計時，如何保留更大的彈性與更豐富的發展空間，使得未來系統需求改變，或是整合其他後設資料標準時，比較容易維護及更新，將是重要的課題。

二、系統建置流程

基本上，數位典藏系統的建置方法與傳統系統建置方法並無太大的差異，同樣是經由規劃、分析、設計、開發四個步驟進行。「規劃」—系統建置前，必須儘可能從各方的需求著手進行通盤的規劃；「分析、設計」—透過完整的系統分析方法，分階段落實系統規格中的設計，以避免造成典藏系統功能開發的重複、衝突或不足等情況的發生；「開發」—經過完整的資訊系統規劃與分析，並整合領域知識專家的典藏需求與應用需求，分階段進行系統與相關的應用功能的開發。按照上述步驟，能建置出符合各項功能需求與目標的系統。

然而對數位典藏而言，除了著重在系統面的考量外，另一個重點就是後設資料的規劃設計。因此要建置一個數位典藏系統，必須經過系統規劃、後設資料分析、系統分析設計與系統開發等不同的工作程序。其流程如圖 3-1：

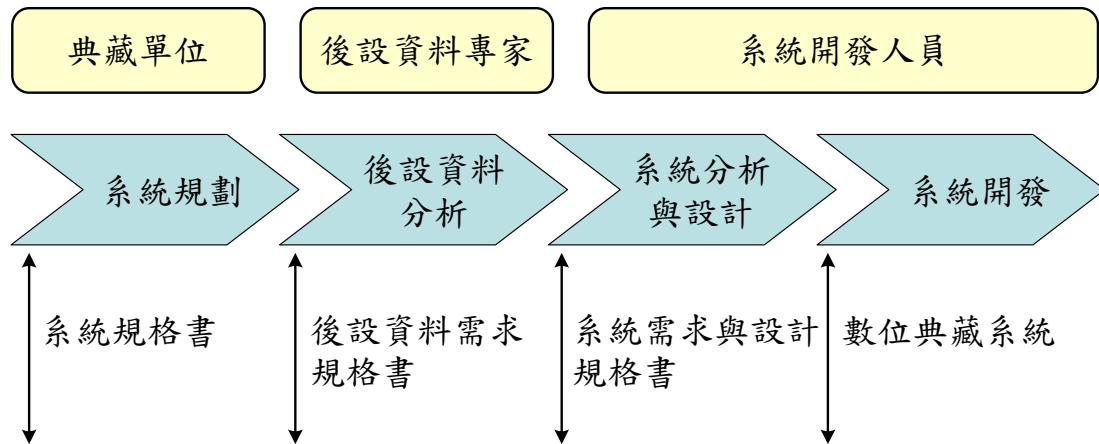


圖 3-1：數位典藏系統建置流程

- 1.系統規劃：由典藏單位針對本身典藏需求，規劃出合適的系統規格與功能說明，並且對於各項技術需求進行初步可行性評估，產出系統規格書。
- 2.後設資料分析：由後設資料專家訪談典藏單位的內容專家，充分了解典藏資料的內容與該單位的需求後，設計出一套合適的後設資料結構與屬性。最後透過製作後設資料需求規格書，以確認後設資料需求，並作為典藏單位、後設資料專家、系統開發人員三方面的溝通橋樑，促使不同專業領域達成相同共識(Chen, Chen, & Lin, 2003)。
- 3.系統分析與設計：為了確保典藏單位的需求，能被完整的呈現於數位典藏系統中，系統開發人員透過與典藏單位溝通，並分析使用者的系統功能性需求後，再進行系統的流程與功能設計。最後利用系統需求規格書及設計規格書來描述系統內容，讓典藏單位能快速了解系統全貌，以確保每一個程序的正确性，避免系統完成後與使用者有認知上的落差。
- 4.系統開發：系統開發人員經過專業的技术與知識訓練後，依照系統分析文件中的規格，利用各種資訊工具開發出系統功能，並且整合相關的資訊技術，如：資料管理（包含著錄、查詢、維護）、資料搜尋（包含影像比對、全文檢索、網頁搜尋、資料庫查詢）及瀏覽（包含分群、分類）等相關技術，以建構出一個完整的數位典藏系統。

三、系統規格分析之考量

數位典藏系統建置的每個過程，都需要花費相當多的人力與時間，並且需由不同角色的專業人員，透過分工方式針對系統進行客製化處理。尤其典藏單位、後設資料專家及系統開發人員三方面，需要密切溝通並分工合作，以減少可能產生人為錯誤的情況，才能讓有限的人力做更有效的運用。另外，在系統規劃時，必須清楚記載系統規格需求，並一再確認需求是否與典藏單位的想法有所出入；而在分析設計時，則必須著重於整體系統的擴充性、互通性、技術性等方面，以設計出最符合的系統架構。因此，下面幾個核心的規格，是必須特別注意的：

- 1.系統功能說明：針對典藏系統所設計的每一項功能，應詳列各項功能的說明，及其各項輸出、入介面規格，並包含使用者操作流程與使用者介面的呈現方式介紹。
- 2.著錄規格：根據需求訪談資料所彙整的著錄項目需求與規範，必須詳記各個著錄項目的中、英文名稱、意義、實例、著錄規範、修飾語與備註等項目，以作為資料操作人員於資料編修時的依據，與系統設計時的準則。
- 3.資料庫欄位定義：根據後設資料需求規格書中，對於著錄規範與內涵關係架構，所設計出來的資料庫欄位規格，是系統開發人員重要的參考依據。因此資料庫欄位規格，必須詳細包含資料表名稱、欄位名稱、欄位別名、欄位型態、欄位大小與備註等說明。尤其欄位名稱與別名需儘可能與著錄規格相同，並須明確記錄該欄位是否需索引、是否為主鍵(Primary Key)或外來鍵(Foreign key)等資訊。
- 4.相關技術需求：典藏系統建置過程中，需要整合各項資訊與技術，因此對於關鍵性的技術議題，應詳細提供確認、評估及選擇技術解決方案的結果，以符合系統發展之目標。

第二節 系統分析

系統分析是典藏系統建置過程中必經的階段，系統的成功與否，掌握在分析工作是否完整且徹底。系統分析的目的，在將使用者的需求以直接、具體、有系統的方式正確且清楚的記錄下來。因此大部份的關鍵因素取決於系統分析人員的訓練與經驗，有經驗的分析人員能夠快速的掌握使用者的需求，減少系統開發時的錯誤與時間的浪費。有鑑於此，本書於下將介紹系統分析應該著重的項目，與一般考量的因素，讓讀者能夠更容易掌握系統分析技巧。

一、使用者分析

數位典藏系統的主要功能，在建置一個完整的數位化環境，並提供各種典藏工作流程所需之資訊服務。對於每個典藏工作，可依照功能需求以及使用模式，對應到不同的使用者族群，如圖 3-2。因此在進行系統分析時，必須了解系統可能的目標使用者（Target users）。目標使用者不同，典藏系統的定位及設計方式亦不同。故透過使用者分析，以清楚掌握典藏系統未來的目標使用者、使用習慣、使用動機等因素，並針對不同的目標使用者，設計出合適的系統功能及操作流程，如此才能為使用者提供豐富、完整的系統服務。一般而言，典藏系統的服務對象分為前端使用者及後端管理者。前端使用者，又可依照使用程度的不同，再區分為一般使用者及進階使用者；而後端管理者，又可依照管理範圍的不同，再區分為內容管理者與系統管理者。以下將對每個角色作說明，並論及其對系統的可能需求：

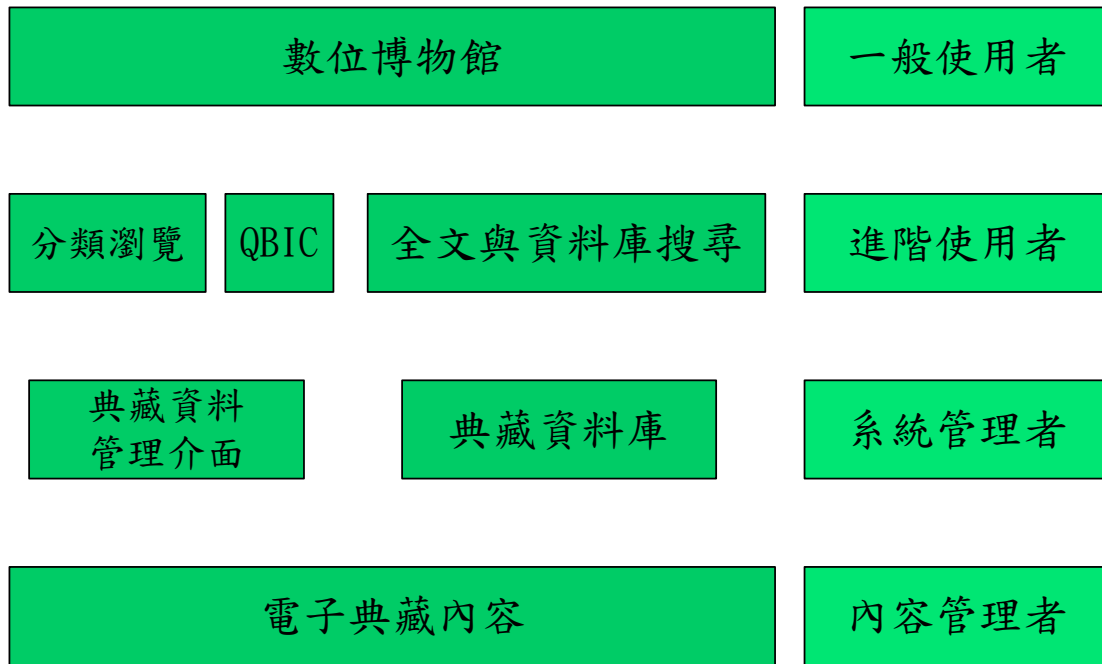


圖 3-2：數位典藏環境架構（范紀文，民 91）

- (一) **一般使用者**：這類人員使用數位典藏系統的主要目的，在於瀏覽或搜尋一般性的資訊，以充實自我。因此在系統設計時，需要一個簡單、一致的檢索與索引介面，以協助取得相關資料。
- (二) **進階使用者**：主要是從事典藏相關工作的人員，或是以特定典藏資料進行更深一層研究的專家學者及研究人員。此類使用者，除了要更深入了解系統功能外，也需要更準確的資訊搜尋工具。
- (三) **內容管理者**：主要負責典藏品資料的整理工作。在系統建置完成後，負責著錄後設資料內容，並校對資料的正確性。除了需要經常性地更新資料外，也掌控了資料對外開放的時機和範圍。
- (四) **系統管理者**：為負責典藏系統設計、製作、維護、與管理的資訊技術人員，不但熟悉系統整體的功能操作，並能使用相關工具技術，來維持系統的正常運作，與管理各項工作流程。

二、系統需求分析

不同典藏單位的典藏系統需求差異甚大，即便是相同性質的典藏內容，亦會因研究目的、觀點的不同，而有不同的系統需求。也就是說，每個典藏單位有其獨立的系統開發工作，且必須依照實際典藏單位的需求，才能設計出符合典藏單位的數位典藏系統。因此系統需求分析，便是透過與典藏單位溝通，以了解典藏

系統的需求，並將系統需求轉換為系統開發人員所了解的技術規格。

從系統需求分析中，可以了解典藏系統可能包含的功能，而系統開發者對領域知識與系統使用者需求的瞭解程度，對整體典藏系統的發展成功與否，影響甚鉅。系統分析人員對該領域的瞭解程度愈深，則與典藏單位的溝通就愈容易；愈瞭解典藏單位的典藏內容，對典藏系統的需求掌握亦愈正確。系統開發者在瞭解系統需求的過程中，必須掌握以下的重點（范紀文、何建明、李德財，民 92）：

- (一) **領域知識的吸收與快速擷取**：充分了解該領域的基本運作、思考模式與語彙，有助於分析工作的進行。從分析過程中，亦可透過與典藏單位的互動，吸收相關的知識內涵，並轉換成對系統發展有用的知識。
- (二) **取得典藏單位的信任**：系統的發展有賴典藏單位相關人員的配合，以對等的心態與本身的專業，幫助典藏單位解決其所面臨的問題，會有助於整體工作的順利進行。
- (三) **充分溝通與了解**：透過再三的溝通與說明，探討典藏單位的需求，讓雙方充分了解彼此的想法後，再作判斷或提供建議，有助於取得彼此的信任，降低雙方的隔閡。
- (四) **掌握典藏內容的內涵**：透過對領域知識與典藏單位的了解，掌握典藏品間的關係及其知識架構。根據該知識架構與典藏品間的關係分析，了解整體典藏計畫所包含的範圍、系統功能規模及著錄項目。

需求分析階段，是系統開發時最重要也最複雜的階段，必須經過綿密的思考及溝通，才能產生較完整的需求。這也依賴豐富的經驗與專業素養，以及在分析的過程中，儘可能了解典藏單位的真正需求，從中歸納出典藏單位的作業流程、系統功能及管理方式等項目。

三、軟體需求規格書

為了將系統分析的結果，直接、具體且清楚的向系統開發人員表達，需將分析出來的需求，用系統化及文件化的方式記錄下來。一般常用的方法，便是撰寫軟體需求規格書（Software Requirements Specification，簡稱SRS）。SRS是用來描述軟體產品之最終使用者的期望需求，它可以用來評估軟體開發的成果，是否達到當初釐定之使用者需求規格，並作為未來測試及驗收的依據。因此在系統分析後，可將需求分析的結果，以文字及圖形的方式說明，撰寫出一份容易了解、需求定義明確的軟體需求規格書，作為系統分析師與使用者溝通之媒介。一份完整

的SRS，一般會包含以下重要項目（林彥君、黃建中、王祥安，民94）：

目錄

目錄列出文件各章節之名稱及頁碼。

1. 前言

1.1. 目的

本節說明文件目的、計畫名稱、委託人及預期的讀者。

1.2. 著錄範圍

本節描述著錄範圍並簡介系統，包含以下項目：

- (1) 系統概述：包含系統名稱、系統建置目的並簡介系統功能。
- (2) 著錄範圍：說明典藏文物特性、出土地與系統資料來源。

1.3. 定義與縮寫符號

本節定義出現於SRS之重要定義、縮寫符號與簡稱。

1.4. 參考資料

本節列出SRS所參考引用之資料清單，並說明各資料之作者、版本、出版日期及超連結等相關資訊，在數位典藏SRS中，常見的參考資料包含系統之後設資料需求規格書及相關系統之軟體需求規格書。

1.5. 綜覽

本節說明文件組織及後續各章節內容。

2. 整體概述

2.1. 產品面面觀

本節說明系統架構及操作介面等相關資訊，包含以下項目：

- (1) 系統架構說明及系統架構圖：說明系統架構，並附上系統架構圖。
- (2) 軟體環境：說明系統所處之軟體環境，如以網頁為操作介面，使用者可透過網際網路及瀏覽器操作系統所有功能。
- (3) 人機介面：說明系統操作介面相關資訊，如螢幕解析度、圖形設計及網頁設計之原則。
- (4) 與外部系統或專業之關係：描述與其它外部系統之整合情形。

2.2. 產品功能

本節以文字及各種圖示簡介系統功能，包含各系統功能的目的、作用及系統功能之間的關係。

2.3. 用戶種類特性

本節說明系統用戶總類及特性，包含負責的工作項目及擁有之功能權限等資訊。

2.4. 限制

本節描述限制開發者之項目，包含管理方法、硬體限制、高階語言需求、系統保護機制。

2.5. 假設與相依性

本節列出會影響軟體需求之因素，如特殊的作業系統。

2.6. 未來需求

本節說明可延後至未來才製作的需求。建議應考量實際開發情況，評估加入本節的必要。

3. 分項功能需求

3.1. 數位化規格

本節描述特殊的典藏文物儲存格式，包含圖片的解析度及大小。

3.2. 功能性需求

本節描述軟體各功能之細部資訊，包含各功能的操作流程及細部設定，並以使用案例圖(use case diagrams)呈現各功能與用戶之間的關係。

3.3. 效能需求

本節說明靜態與動態之軟體效能需求，包含可支援之終端機個數、同時操作人數及在固定時間處理的資料數及型態。

3.4. 資料庫的邏輯需求

本節說明要放入資料庫之資訊的邏輯需求，較細節的資料庫設計資訊則在軟體設計規格書說明。

3.5. 設計限制

本節說明由網路通訊協定、其他標準及硬體限制所引發的設計限制。

3.6. 軟體系統屬性

本節描述軟體應具有之屬性及在使用操作上及資料保護的要求，包含可靠度、可利用性(availability)、安全性、可維護性、可攜性。

附錄

附錄包含輸入/輸出的範例格式、讀者審視後的結果及有助於讀者了解的相關背景資料，如需求回溯表、工作時程規劃表、會議記錄及著錄範例等。

中英文索引

中英文索引可幫助讀者尋找文件中出現之重要名詞及特殊的縮寫。



第三節 系統設計

系統設計是一個非常專業的工作，不僅系統人員要對系統規格有充分的了解，本身也要具備一定的技術背景。尤其在系統範圍的掌握上，必須要以宏觀的角度切入，才能瞭解哪一個系統的哪一個流程的哪一個操作會有哪些使用者來使用，以及是否需要與其他系統整合等。如此才能清楚、明確地掌握各子系統之間，如何制訂標準的溝通規格、如何傳遞訊息，以及要用什麼樣的技術來開發。數位典藏系統的設計，通常必須考量典藏單位的實際使用情境，並根據系統需求設計出適合的系統架構與功能，絕對不能單從技術的觀點以偏概全。本節將針對於系統架構與功能的設計部份加以說明。

一、系統架構設計

一般的數位典藏系統，基本上要滿足前端檢索呈現，與後端資料管理兩個部份的需求。前者提供一般使用者檢索與瀏覽典藏資料的介面，後者為典藏單位管理後設資料的管理控制介面。系統設計針對以上兩個需求，從典藏導向與使用導向的角度，設計出典藏系統所應具備的相關子系統，如下圖3-3（黃國倫、李家豪、何建明，民92）。

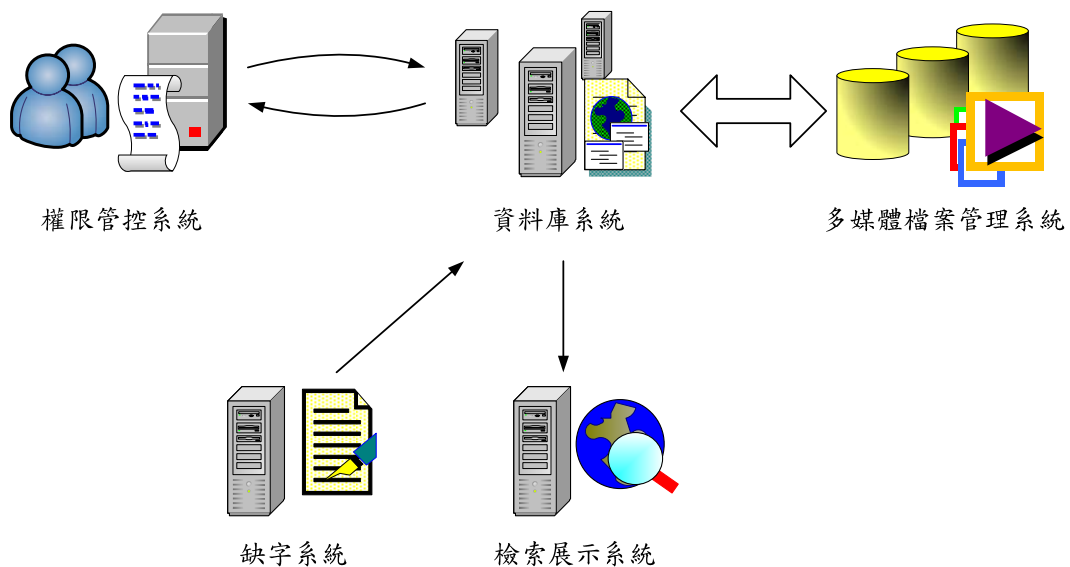


圖 3-3：數位典藏系統架構

(一) 資料庫系統

數位典藏的目的，是希望透過資訊技術以完整保存典藏文物的資料內涵。在典藏系統中，必須清楚且完整的呈現出藏品應有的性質與特性，以及與其他藏品間的相互關係。因此，將文字、數值、圖表、影像等數位化資料及後設資料，經過有系統的整理後，建置為資料庫，除了方便資料管理外，也可提供各式資訊的整合應用。

由於不同廠商所提供的資料庫工具，在功能、規格與資料存取的效率上有所差異，因此資料庫的建置是有相當的難度。在資料庫的介面設計上，建議採用 WWW 架構，透過網頁操作，可提供多人同時線上著錄與檢索資料，並能夠達到跨使用者平臺的優點；不要像過去的資料庫，應用程式和資料操作介面往往綁在一起，導致程式散佈及更新上的困難。

另外，在進行資料庫設計時，除了考量到後設資料的管理與檢索功能之外，也要考量如何將資料庫的資料與多媒體的資料互相整合連結，以便日後展示使用時，能夠呈現出最完整的典藏資訊。

(二) 多媒體檔案管理系統

所謂的數位化資料，不侷限於文字內容，往往圖片、聲音、影像等多媒體檔案亦會加以數位化，進行保存。因此，數位典藏系統對多媒體資料的管理工作便顯重要。一般市面上的資料庫系統僅能對文字性的後設資料進行儲存、管理，多媒體的資料雖然也能夠儲存，但卻無法有效管理，以執行多媒體檢索、資料轉檔、線上展示等。因此完整的數位典藏系統，需要擁有多媒體檔案管理功能，以負責檔案保存與管理工作（鍾子帆、何建明，民91）。這個部份將在第五章進行詳細介紹。

(三) 檢索展示系統

數位典藏的另一個目的，是為了提高典藏資料普及性，讓使用者可以透過網路檢索及瀏覽典藏資料。由於典藏系統中保存大批的數位化資料，必須透過有效的查詢方式，使用者才能從眾多資料中找尋出自己所需要的內容。因此在典藏系統中，應該建置一個有效的檢索展示系統，不只提供適合使用者的檢索或瀏覽方式，在呈現上，也應該結合多媒體圖片、影像或全文檔的內容，以完整展示典藏資料。另外，由於各領域的使用者習慣與典藏資料特性有所不同，對於典藏資料展示的方式、介面的操作流程、設計的風格等，應與典藏單位充分溝通。尤其在使用者介面設計上，若能進行使用行為評估並檢討改進，將可使系統之設計更加完善。這個部份將在第四章進行詳細介紹。

(四) 缺字系統

在數位化過程中，往往會面臨某些古字或罕用字，無法於系統中輸入、顯示與查詢，這就是缺字問題。尤其在人名、地名部份，往往有相當嚴重的缺字問題，若不能有效的克服，整個系統功能勢必受到限制。所以數位典藏應該導入缺字的解決方案，將缺字系統整合至典藏系統之中，以順利地進行資料數位化工作（黃國倫、蕭人豪、李家豪、陳心渝，民93）。

（五）權限控管系統

對典藏系統而言，除了要具備足夠的功能與資料之外，不同角色的使用者應該擁有不同的權限，也是數位典藏系統必備的功能之一。因此，權限控管系統，除了負責維護使用者帳號外，也應提供每個使用者權限控制設定，例如：系統管理權限、資料著錄權限、代碼維護權限、資料檢索權限等。另外，也應該紀錄使用者操作行為，例如：何時操作典藏系統、修改那一筆資料、查詢什麼關鍵字等，作為人員稽核或是系統評估的參考依據。

二、系統功能設計

在上述架構中，典藏系統是由各子系統所組成。因此在功能上，除了具備上述子系統所包括的基礎功能外，不同典藏單位也會依據本身作業或管理上之目的，而發展更多系統功能，以下是幾個常見的系統功能：

（一）庫房管理功能

任何典藏單位皆會面臨典藏管理的工作流程問題，而數位化典藏資料管理往往與實體的典藏管理是密不可分的。以器物類典藏品為例，後設資料管理往往會與其內部的作業管控、研究鑑定、調件取用、修繕、展覽出版等作業相關。這些庫房管理功能，雖然不屬於數位典藏的核心功能，但卻是數位典藏工作流程的重要需求。因此在典藏系統中，除了基本的典藏品資料建置與維護功能外，也應將庫房管理的功能與數位典藏的資料管理整合，並配合內部作業流程，建構出完整的庫房管理功能，以輔助原本的典藏管理工作。

（二）系統安全機制

系統的安全機制，可分成兩個方面的能力來探討，第一是系統備援與復原能力，二是系統安全與防護能力。一般影響系統穩定度的因素，不論是人為破壞或自然災難都造成系統無法正常運作。因此平時應規劃系統復原程序，包括備援系統的設置、損害範圍隔離策略(Defect Isolation)制定、補強能力的評估等各項作業，以降低系統停機所受到的損失。其中人為破壞部份，可能是系統遭受攻擊，如：網站遭受到的阻絕癱瘓攻擊（Denial of Service Attack，簡稱DoS），在短時間內，有大量的網路連線，而造成網站對外服務中斷。也有可能是因為軟硬體系

統本身的可能缺失，或使用者管理的不當，而遭到入侵或破壞，常見的情形包括帳號蒐集攻擊法 (Directory Harvest Attacks, 簡稱DHA)，以及利用作業系統的漏洞，以取得系統管理者權限，而規避正常的權限檢查程序等。因此在系統服務安全策略的規劃上，除了對系統不當使用頻率的監控外 (包括無意義的資料查詢、過度頻繁的連線需求)，也應著重於安全稽核作業的完整性，並配合適當的存取控制，如防火牆安全設定、人員權限查核等，以防止惡意攻擊。

(三) 數位權利管理機制

就數位典藏目的而言，除了進行典藏資料的保存外，也會發展數位資料的應用。因此當典藏系統建置完成後，最重要的工作便是對外開放使用。然而數位化的資料往往輕易就能複製，任何經過授權的使用者，可能透過正常的步驟存取系統中典藏資料，並將取得的資料經過複製後，再轉手提供給他人，而造成典藏單位在權益上的損失。關於這一方面的問題，需要建立數位權利管理機制來控管資料完整性、私密性與使用權利，並提供數位化版權識別、防止非授權之使用及資料竄改、使用及擷取契約機制、記帳及付費機制、測量及監控著作使用情形等功能，以確保數位資料使用的合法性。這個部份將在第七章進行詳細介紹。

(四) 系統擴充性及負載平衡機制

在系統設計時，必須考量系統服務的品質。所謂系統服務的品質，包括同一時間內系統的最多使用人數和最大的處理資料量，這也代表著系統所能承受的流量及負載。若是在網站上只提供文字資料或低解析度的圖片，對系統流量及負載或許不成問題。但是，如果開放高品質圖片或影片檔存取，有可能造成系統負荷不了，甚至於當機的狀況。因此典藏系統建置時，在評估過服務品質的需求後，可以加入擴充性及負載平衡機制的功能設計。即使未來面對流量的成長，只要透過系統設備的擴充，就能夠維持合理的服務品質。

(五) 資料備份與異地備援機制

儲存及備份機制為資訊系統永續經營的基礎。它對於資料的長久保存與資料價值之發揚與傳佈，舉足輕重、不容誤失。因此需要有最經濟、快速之方式，提供系統中的資料有效、穩定的儲存與備份方案，以提昇資料安全保障。

不同於資料備份，異地備援不僅僅拷貝資料而已，而是在另一地點建置一套相似的備援系統，並透過網路將資料傳送到備援系統。一旦系統發生突發狀況無法運作時，備援系統會取代運作任務，使系統的停機時間降至最低，並確保資料的完整，讓系統能以最短的時間恢復正常運作。故建立完善的異地備援機制，是面臨重大危難事件之後，能否快速順利復原之關鍵，由美國 911 事件即可獲得明證。

(六) 典藏資料交換

典藏資料的分享必須架構在相同的著錄規範下，但因不同典藏單位間彼此數位化的目的與觀點不同，著錄規範亦會有所不同，如此將造成彼此間資料無法流通的嚴重問題。典藏資料交換機制的建立，便是為解決典藏資料互通性的問題，以達成分享彼此研究成果的目的。

第四節 資料庫開發

資料庫是典藏系統重要的核心，藉由資料庫工具提供的儲存、搜尋以及管理功能，可以提供典藏系統不同方面之應用。然而對數位典藏系統的資料庫而言，最重要的功能是後設資料管理能力，包含資料的變動（新增、修改與刪除）或是應用（查詢）等。因此在資料庫開發過程中，典藏品的後設資料分析是很重要的一環。其分析結果會影響到典藏系統的完整性，以及資料庫中典藏資料著錄及檢索的精確度，更會影響後續的資料交換與加值利用的可能性。

在資料庫設計上，大多是使用關聯式資料庫工具。其基本概念請參考第四章。這樣的方式可以很快速的處理固定或穩定的資料結構，而且目前市面上有相當多的產品可供選擇。然而，在不同的領域所建立的後設資料格式常常有相當大的差異，為了考量不同的使用者需求，資料結構往往複雜且經常變動，因此關聯式資料庫會受到很大的限制。例如，在資料庫欄位設計時，要將一個完整的後設資料拆解成數個關聯式資料表格，是需要複雜的分析工作；或是當資料儲存在許多資料表格時，如何將其合併成一個完整的後設資料，也需要複雜的程式設計工作。即使系統正式運作後，因需求改變而變更資料結構，更是一件困難的工作。

針對這個問題，有些資料庫在設計上會採用XML資料庫工具，由於XML具有標籤，可用來儲存結構化文字資料，又因具有高彈性、文件結構容易修改的特性，常被用來儲存資料結構複雜或變異性高的資料內容，也被廣泛應用在資料交換格式上。使用XML資料庫工具，可讓後設資料的結構與內容更具彈性，也可容易的匯出XML格式的文件，與其他系統交換。因此於資料庫的開發方法上，本節將特別針對XML資料庫的設計加以介紹。

一、後設資料規格分析

對典藏系統而言，資料庫主要的功能在管理典藏資料，而數位典藏所涵蓋的典藏資料除了數位化的多媒體資料外，也包含後設資料的內容。因此在著手資料庫開發前，首要的工作便是後設資料規格的分析。透過後設資料分析才能了解後

設資料欄位的意義、屬性設定以及資料庫的操作介面的設計和資料著錄的方法等重要資訊，以充分掌握資料庫與後設資料的關係。

為了清楚後設資料規格，必須詳讀後設資料需求規格書中的各項內容，如：後設資料結構、元素屬性、著錄範例、控制詞彙或代碼清單、相關標準的比對與系統功能需求等，皆是資料庫設計時的參考依據。尤其下列幾個重要的後設資料特性，更是系統分析人員必須考量到的：

- (一) **多表單結構**：傳統的表單分析方法會在一份表格中揭露所有欄位資料，造成表格極為龐大、管理不易。然而在後設資料分析時，會依據典藏品本身的性質切割成許多層次，如表 3-1。因此一筆完整的典藏品後設資料會包含多個表單結構，如圖 3-4。延續上述概念，當資料庫設計時，也會將每個層次的資料，依性質切割成許多的資料表格，並且在每個表格中設計連結欄位。在後設資料著錄的同時，就會透過程式將表格互相串連，將完整的後設資料儲存起來。

表 3-1：器物後設資料著錄範例

元素中文名稱		實例
Element	Subelement	
器物層次		
品名	中文品名	伯三 小隔
	英文品名	Po shih li
	原品名	古銅隔
時代	起	1100 B.C.
	迄	1000 B.C.
保存狀況		良好
出土地		
展覽層次		
展覽編號		
展覽名稱	中文	文存周金~西周金文特展
	英文	A Special Exhibition of Bronze Inscriptions from the Western Chou
展覽時間	起	2001/06/30
	迄	
策展者		游國慶
圖檔層次		
圖檔檔名		C1A001658N000000000AA
說明		
開放限制		開放

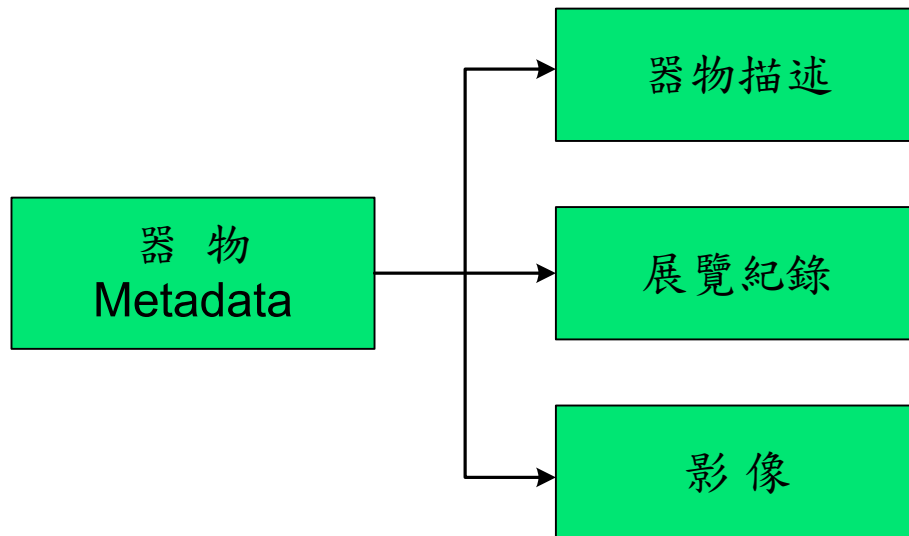


圖 3-4：器物後設資料內涵關係圖

(二) **多值欄位結構**：後設資料中，某些著錄項目會有多值著錄需求，亦即在一個欄位中會需要著錄一個以上的資料內容。此種結構的資料需要額外的處理方式，以方便使用者著錄多筆資料內容。根據歸納分析的結果，一般多值欄位結構又可細分成四種不同的建置形式（陳亞寧、陳淑君、沈漢聰、鍾豐謙，民 91），如下所列：

a. 一個多值欄位

例如：【故宮書畫數位典藏子計畫】

項目名稱	多值欄位	著錄一	著錄二
技法	◎	水墨	皴法 (披麻皴)

b. 一組多值欄位

例如：【故宮器物數位典藏子計畫】

項目名稱	多值欄位	著錄一	著錄二
品名	類別	◎	中文品名
	品名		玉鳥
			原品名
			商舊玉蟻形珮

c. 一組多值欄位裡有一個或一個以上多值欄位

例如：【近史所檔案館外交經濟重要檔案數位典藏計畫】

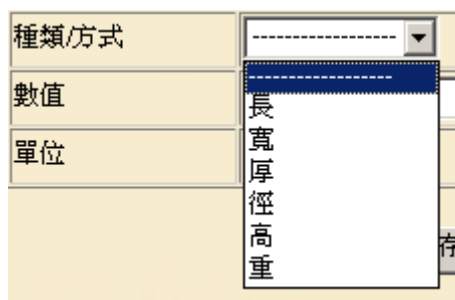
項目名稱	多值欄位
別名權威控制	機關權威
	職銜權威
	人名權威

d. 一組多值欄位裡有一組或一組以上多值欄位

例如：【國史館典藏國家檔案與總統文物數位化中程計畫】

項目名稱	多值欄位	
檔案附屬層級	系列號	
	系列名	
	副系列	副系列號
		副系列名
	宗	宗號
		宗名

(三) 代碼選單欄位：在後設資料需求中，部份欄位並不同於一般的文字欄位，可以著錄任意資料，而是必須自限定的選項中加以挑選，一般稱為「代碼」。代碼欄位在建檔時往往作詞彙控制之用，並且以下拉式表單呈現。因此在資料的著錄介面上，可以使用點選的方式加快資料輸入的速度，也可避免資料輸入錯誤或超出代碼範圍的狀況發生，如圖 3-5。在資料庫設計上，除了應考量到系統著錄時，對於代碼資料的儲存方法外，也必須規劃額外的代碼表格，以提供維護代碼選項的功能。



書寫工具	<input type="checkbox"/> 毛筆書寫 <input type="checkbox"/> 刀刻 <input type="checkbox"/> 其他
書寫顏料	<input type="checkbox"/> 墨 <input type="checkbox"/> 硃砂 <input type="checkbox"/> 其他
書體	<input type="checkbox"/> 隸書 <input type="checkbox"/> 草書 <input type="checkbox"/> 其他

圖 3-5：兩種代碼呈現介面

(四) 典藏品關係結構：在典藏品與典藏品之間，往往存在某種關係，這種關係在後設資料分析時，會利用層級關係圖與群組關係圖來表示，如圖 3-6 及圖 3-7。為了表現出這種關係，在資料庫設計時就必須建立額外的關係表格，以記錄典藏品間的關係。不同的關係結構，設計上也會有所不同。有時候，某些特殊的關係描述可能會比典藏品本身的後設資料還要複雜。

原件(典藏品)之藏品單元層級關係

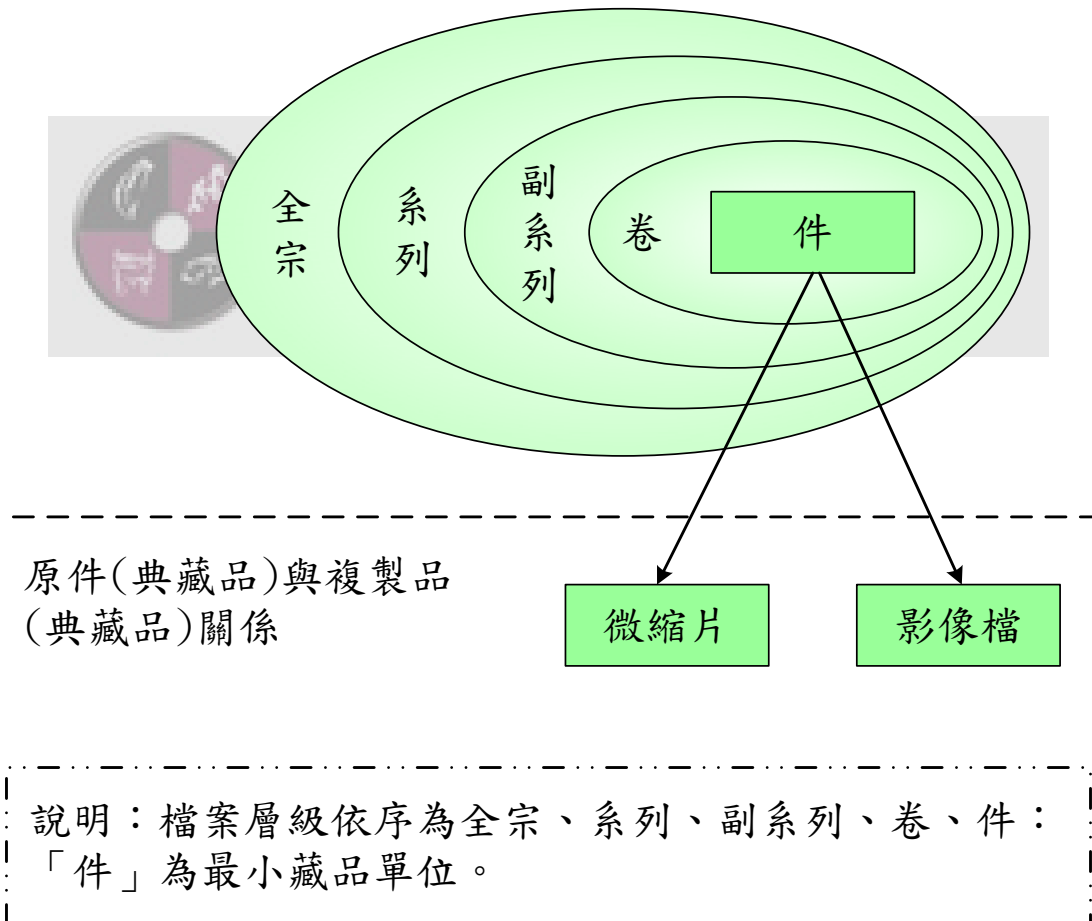
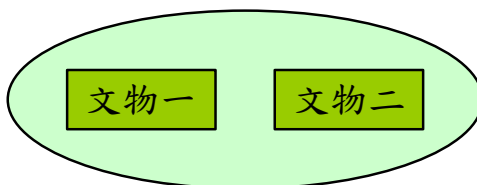


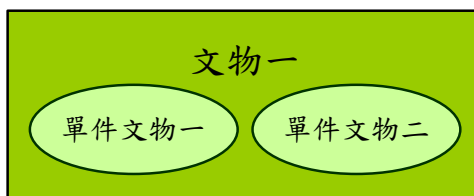
圖 3-6：檔案藏品層級關係圖

群組關係一：組件



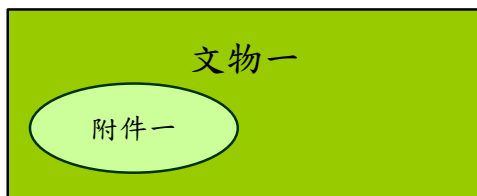
說明：最小之藏品單元(unit)為文物，兩文物會組合為有意義之一組文物組(即為組件)，最小之藏品單元為組件皆各需著錄一筆紀錄。

群組關係二：合併



說明：最小之藏品單元(unit)--文物可能為一組文物集合(即為合集)，只就藏品單元(unit)著錄。

群組關係三：附件



說明：最小之藏品單元(unit)--文物可能包含不著錄之附件關係。

圖 3-7：藏品群組關係圖

(五) 欄位屬性：除了上述的幾種結構特性外，後設資料的設計上，也有幾個重要的欄位屬性會影響到資料庫功能設計，說明如下：

1. 必填欄位：是指在建立後設資料紀錄時，必須填寫的重要基本欄位，以確保資料的完整性。故應在系統中建立管控機制，以避免不完整的紀錄被存檔。
2. 預設值：某些欄位會有預設著錄的資料值，例如：若某一類典藏品的「保存狀況」欄位通常為「現況良好」，則可針對此欄位設定預設值為「現狀良

好”。當資料著錄時，會自動填入“現況良好”，或依實際情況加以修改，以加快資料著錄。

3. 唯一值欄位：唯一值的設計目的在確保典藏資料不重覆，因此會以唯一值欄位做為檢查判斷之用。每當資料新增或更新時，必須自動檢查資料庫中該欄位所有的資料，即時判斷是否重覆，以提醒著錄人員。

二、XML 的後設資料規格

由於XML具有可自訂標籤 (Tag)、可擴充性、易於程式處理及可彈性地改變文件結構與內容的特性，因此常被用來呈現結構複雜的資料。而在數位典藏中，為了完整的保存典藏資料的特色，在Metadata的設計上往往採取複雜的資料結構，而且資料格式也會經常改變。若能以XML文件作為Metadata的儲存單位，對於Metadata的架構及內容會有較高的彈性。XML也常被用來表示Metadata規格，如應用在描述數位化資源的Dublin Core (如下圖3-8) 及描述學習物件的 Learning Object Metadata等Metadata標準，皆訂出以XML 呈現Metadata的規範。

```
<?xml version="1.0"?>
<metadata
  xmlns="http://example.org/myapp/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://example.org/myapp/ http://example.org/myapp/schema.xsd"
  xmlns:dc="http://purl.org/dc/elements/1.1/">

  <dc:title>
    UKOLN
  </dc:title>
  <dc:description>
    UKOLN is a national focus of expertise in digital information
    management. It provides policy, research and awareness services
    to the UK library, information and cultural heritage communities.
    UKOLN is based at the University of Bath.
  </dc:description>
  <dc:publisher>
    UKOLN, University of Bath
  </dc:publisher>
  <dc:identifier>
    http://www.ukoln.ac.uk/
  </dc:identifier>
</metadata>
```

圖 3-8：以 XML 呈現 Dublin Core 的資料

要以XML文件呈現後設資料，主要是參考後設資料需求規格表。以漢代簡牘的後設資料為例，在表3-2的欄位需求表中，英文名稱欄位用來說明後設資料各元素的英文名稱。資料型態欄位用來說明著錄資料時的資料型態，如Varchar 表示該元素在典藏系統中是以文字型態的方式來輸入資料。大小欄位用來說明該元

素在典藏系統資料庫中所需之最大空間，以字元為單位。

表 3-2：漢代簡牘後設資料欄位需求表部份內容

項目名稱		英文名稱	資料型態	大小	
基本資料					
識別資料	遺物登錄號	Identifier	Registered Number	Varchar	30
	簡號		Pek. Coll.	Varchar	200

表3-3為欄位屬性表部份內容，其中，必填欄位標示「*」，表示著錄資料時該元素之值不能為空值。屬性欄位標示「唯一」，表示該元素的值在資料庫中是唯一存在的。提供者欄位說明著錄該元素資料時，是由典藏系統自動產生或由填表者自行填入。

表 3-3：漢代簡牘後設資料欄位屬性表部份內容

項目名稱	必填	多值	屬性	提供者
基本資料				
識別資料	遺物登錄號	*		唯一 填表者
	簡號			唯一 填表者

因此，以上述二個Metadata規格所設計出來的XML，將如圖3-9所示。在主題的標籤中，包含一個name 的屬性，其值為「漢代簡牘」，它用來界定後設資料的類別。接著根據前述的表格內容，以標籤代表後設資料規格中的元素（如識別資料、遺物登錄號），並在標籤中以屬性來表示各元素之描述項目（如英文名稱、資料型態等）。

```

<?xml version="1.0" encoding="BIG5" ?>
-<主題 name="漢代簡牘">
  -<基本資料>
    -<識別資料 英文名稱="Identifier" 必填="*" 多值="">
      <遺物登錄號 英文名稱="Registered Number" 資料型態="Varchar" 大小="30" 必填="" 多值="" 屬性="唯一" 提供者="填表者" DC-Element="Identifier"> <遺物登錄號>
      <簡號 英文名稱="Pek. Coll." 資料型態="Varchar" 大小="200" 必填="" 多值="" 屬性="唯一" 提供者="填表者" DC-Element="Identifier"> <簡號>
    </識別資料>
  </基本資料>
</主題>
    
```

圖 3-9：以 XML 呈現漢代簡牘的資料

三、XML 資料庫設計

近年來興起一種專為儲存 XML 文件的原生型 XML 資料庫 (Native XML Database) 工具 (如: Tamino、eXist、Xindice 等)。它是以 XML 文件為基本儲存單位,不需將文件拆解後儲存於多個表格,或預先設定各欄位的資料型態。因此,容易存取整份 XML 文件,並且不會改變其原本的資料結構與內容。除上述優點外,部份工具還提供 XML 查詢 (如: XQuery、XPath)、全文檢索(Full-text searching)、自動索引(Automatic indexing)、Java 程式語言整合(Java binding)等機制,可容易的進行程式的開發與應用。

由於 XML 資料庫對 XML 文件的處理機制較完整,且有開放原始碼的工具可以使用,因此在 XML 資料庫設計上,以原生型 XML 資料庫工具作為後設資料儲存媒體,並採用三層式架構,如圖 3-10。在這架構中,包含了 XML 資料庫、應用程式伺服器和使用者的瀏覽器三個部份。XML 資料庫扮演著資料儲存庫的角色,主要是儲存 XML 格式之後設資料內容。應用程式伺服器扮演著與 XML 資料庫系統及使用者瀏覽器溝通的角色;所有的系統功能都是在此應用程式伺服器上運作。使用者瀏覽器則是使用者的操作介面,因為是以網頁方式呈現,所以使用者無需特別安裝應用程式,即可操作系統上所有的功能。

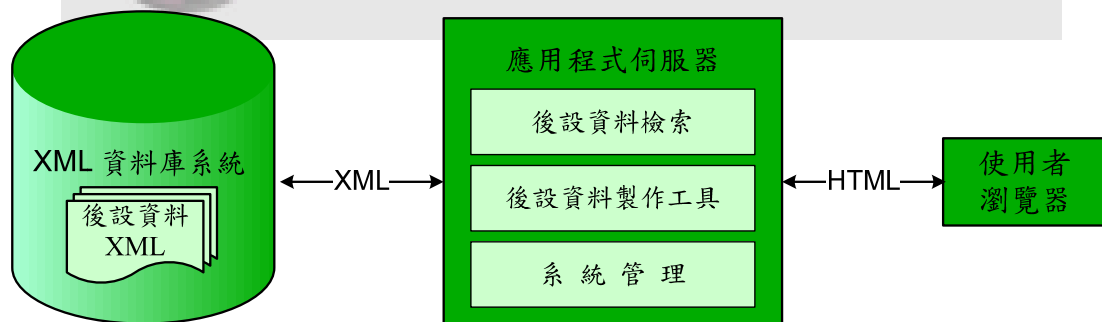


圖 3-10：資料庫系統架構

以XML為基礎的資料庫系統,對於資料的輸入、儲存、查詢、取用、交換與管理都是直接針對XML處理。在後設資料分析完成後,就可以快速的將後設資料規格轉換成XML Schema或XML DTD的資料結構定義。利用XML 結構定義即可快速的產生使用者的著錄介面(如:W3C的XForm標準),經過著錄產出的XML文件就直接儲存於原生型XML資料庫。當需要資料查詢與取用時,透過XPath、XQuery等直接取用資料庫中的整個或部份的XML文件,再透過XSL將XML轉換成一般瀏覽器可支援的HTML格式,使用者就可以看到完整的資料。其系統流程圖,如圖3-11所示。

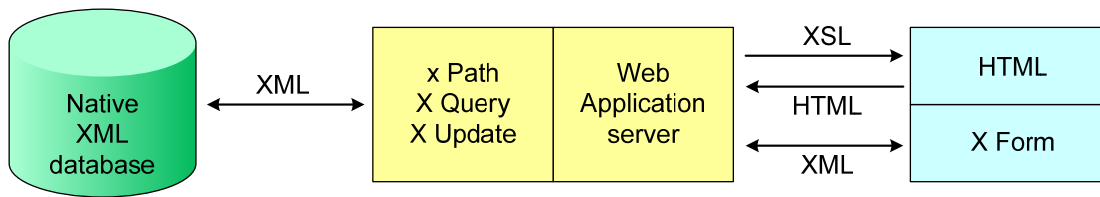


圖 3-11：XML 資料庫流程圖



第五節 使用權限管理

隨著網際網路(Internet)的蓬勃發展，以網頁為基礎(Web-based)的服務也越趨於多樣與完整。網際網路是屬於開放性的環境，即所有的使用者都能夠在網際網路上獲取所需的資訊，但高價值或機密性的資料是不允許被任意存取的，若被輕易且大量的取得，將使著作權人的權利受到損失，因此更突顯權限管理的重要。目前的數位典藏系統，絕大部份建構在網際網路之上。所以如何確保數位典藏的成果能夠開放，並在嚴格管控的環境下被引用與存取，將會是相當重要的課題。

在數位典藏系統中，為了讓多人共同完成典藏資料的管理與維護，所以不同職務的使用者應賦予合適的權限。再者，由於每筆典藏資料具備有獨立與完整性，所以權限控管的範圍也必須能細分至每一筆資料。目前的數位典藏系統大多是以網頁為基礎，每個系統功能可能包含數個網頁的操作流程，所以應將數個網頁置於同一權限控管之下，並以使用者為基礎建立權限管控機制，以解決多使用者與存取權限(Access Right)間的複雜關係。

另外，隨著數位典藏發展越趨完備，使用者往往需要在不同的典藏系統中相互連結操作。為了提供使用者一致的使用環境，在權限架構之中也應該引入單一登入(Single Sign-On，簡稱 SSO)的概念，以避免使用者在不同的典藏系統間切換時，必須重複登入的不便問題。因此為了讓系統開發人員在系統建置時，能夠對使用權限有基本的認識，本節將介紹一般使用權限的重要概念，並針對跨系統單一登入系統深入說明。在本節最後也提出一個符合數位典藏環境的權限管控架構，以作為權限管理功能設計時的參考。

一、使用權限概念

一般常見的使用權限管理方法大致上可分為強制型的使用權管控(Mandatory Access Control，簡稱 MAC)、隨意型的使用權管控(Discretionary Access Control，簡稱 DAC)與角色為基礎的使用權管控(Role-based Access Control，簡稱 RBAC)(黃景彰，民 89)。

(一) 強制型的使用權管控(MAC)：

MAC 指的是由系統管理者統一規範安全政策與資源屬性，且強制實施，使用者無法自行更動，適用於高機密性的單位，例如國防單位。MAC 又是以資訊的機密等級作為執行權管制的依據，因此又稱為以規則為基礎(Rule-based)的執行權管制。

(二) 隨意型的使用權管控(DAC)：

DAC 是由使用者自訂系統操作權限與資源屬性。由於 DAC 是以個體 (使用者) 為單位作為執行權管控的識別, 所以 DAC 又稱為以個體為基礎 (Identity-based) 的執行權管制。

(三) 角色為基礎的使用權管控(RBAC)：

RBAC 是由美國國家標準局 (NIST) 所提出, 且是目前被廣泛討論的一種執行權管控。在 RBAC 中系統權限並不像 DAC 是直接賦予使用者, 而是賦予角色。在企業之中不同的職務就像不同的角色, 依據職務上的需求而給予適當的權限, 再指派使用者至角色上。當然也有可能一位使用者同時被指派到多個角色上, 所以使用者和角色之間, 存在多對多的關係。每個角色則會設定其擁有的權限, 若以系統的功能為例, 權限的設定可以是新增、修改、刪除與查詢等功能。為讓某個角色能完成其所負責的任務, 將會從所有的系統功能中, 授予適當的權限功能, 故角色和權限間也具有多對多的關係。

二、單一簽入系統 (Single Sign-On, 簡稱 SSO)

自主從式架構 (Client Server) 興起, 加上網際網路的盛行, 資料的儲存及運算不再集中於單一主機或平臺上, 這也造成每個系統各自擁有獨立的資料庫及使用使用者驗證機制。此時, 當使用者欲進入任何一套系統, 必須通過該系統之身份驗證 (通常為 Username / Password), 造成使用者的不便及管理上的困難, 故單一簽入系統的解決方案就因應而生。使用者只須利用目前廣泛使用的瀏覽器通過一次身份驗證後, 就可以自由的存取網路上其他的系統服務, 而不必再重複身份驗證的步驟, 如此便可以帶來以下的效益 (廖鴻圖、林金龍, 民 94)：

- (一) 減少多次身份認證所花的時間, 相對的也減少身份認證出錯的可能性。
- (二) 避免使用者同時保有多個認證資訊而造成混亂。
- (三) 管理者可以集中管理使用者擁有的存取權, 並減少維護的時間。

隨著單一簽入系統技術的發展與應用, 在此架構中所定義的成員也越來越多元化, 茲將目前常見的單一簽入系統成員說明如下：

- (一) **使用者 (User)**：又稱用戶端 (Client) 或客戶 (Customer), 其扮演對系統進行資料存取的角色。
- (二) **服務伺服器 (Service Server)**：又稱網站 (Web Site) 或網站主機 (Web

Server) ，其扮演網路服務提供者的角色。

(三) **驗證伺服器 (Authentication Server)**：又稱登入主機 (Login Server) 或驗證主機 (Verification Server) ，其扮演對使用者進行身份驗證，並簽發會議金鑰(Session key)的角色。

常見的單一登入系統實作架構有以下兩種模式 (廖英彥，民 94)：代理人模式(Proxy Model)及重導模式(Inform Model)。

(一) 代理人模式 (Proxy Model)

代理人模式的 SSO 架構中，使用者首先登入到服務伺服器，並提出服務請求 (Service Request) ，服務伺服器則對使用者提出身份驗證請求，使用者將身份驗證所需之相關資訊提供給服務伺服器，此時服務伺服器將身份驗證所需之相關資訊轉送給驗證伺服器驗證，驗證伺服器將驗證結果通知服務伺服器，服務伺服器再依驗證結果決定是否提供使用者相關之服務，如圖 3- 12 所示。

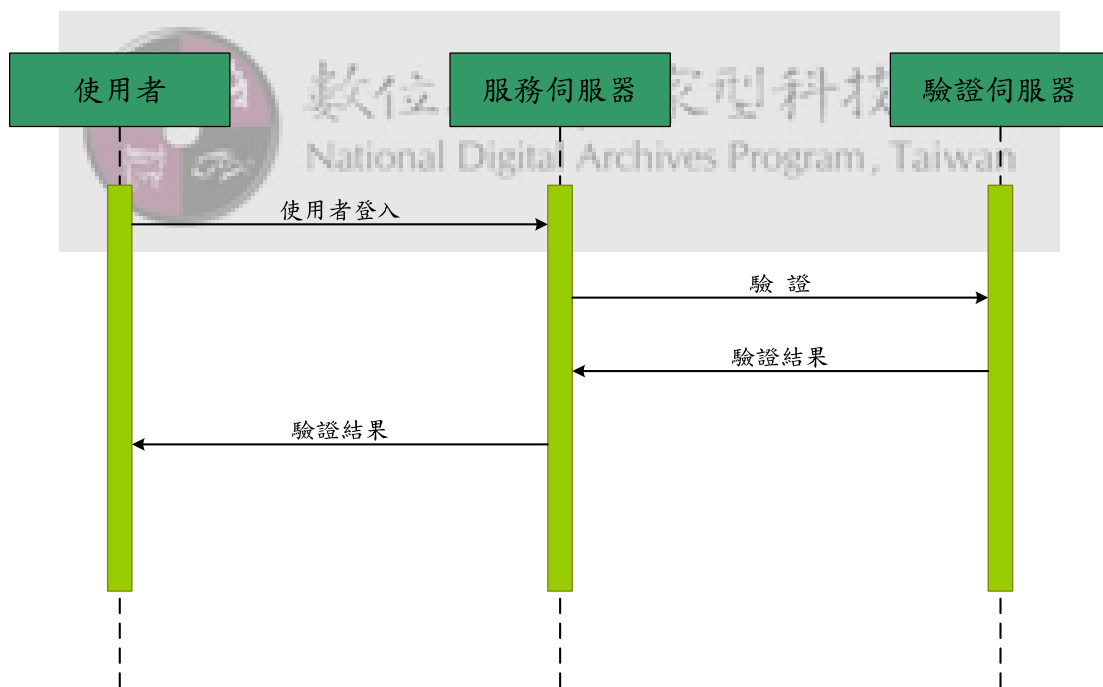
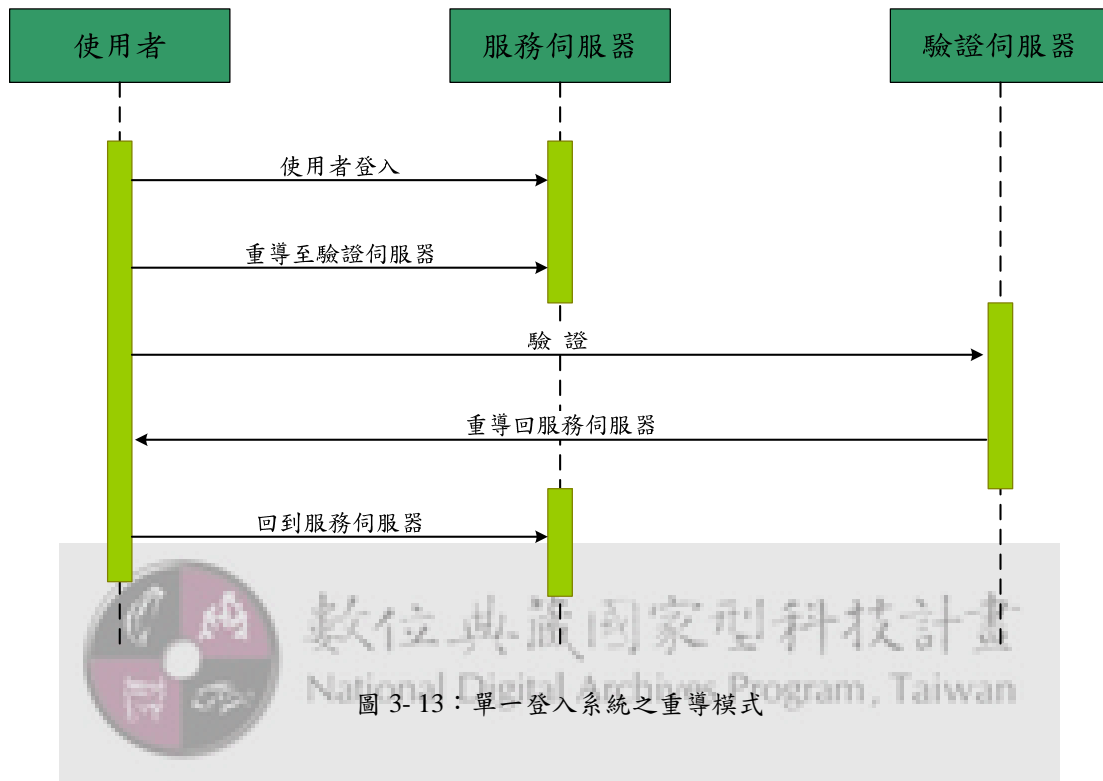


圖 3- 12：單一登入系統之代理人模式

(二) 重導模式 (Inform Model)

重導模式的 SSO 架構中，使用者對服務伺服器提出服務請求，服務伺服器將使用者以 HTTP 重導(HTTP Redirect) 方式，將使用者重導至驗證伺服器進行

身份驗證，若身份驗證成功，驗證伺服器將會簽發會議金鑰予使用者，並將使用者重導回原來服務伺服器所在之網頁，再由服務伺服器驗證會議金鑰之真偽，若會議金鑰驗證無誤，則提供使用者相關服務，如圖 3-13 所示。



不論是代理人模式或是重導模式都各有其優缺點。採用代理人模式在伺服器端的整合會比較複雜，伺服器在處理每一個請求(Request)時必須處理較多狀況，但是在使用者端的設計將會簡單許多；相反的，採用重導模式於用戶端的設計會比較複雜，而伺服器端的設計較代理人模式來得簡單。另外，此兩種模式還有一個比較顯著的差別，在於代理人模式中，使用者只需面對服務伺服器，而驗證伺服器甚至可以部署於內部的私人網域中；而重導模式中，使用者除了面對服務伺服器外，亦必須面對驗證伺服器，而服務伺服器及驗證伺服器間不需要建立網路連線。

三、數位典藏權限管控—範例設計

由於數位典藏系統所提供的服務，是以使用者為導向的資訊平臺，因此系統運作上會有不同角色的人員負責執行不同的系統功能，如：系統管理、資料著錄、資料校對與資料查詢等。若組織的結構有了變動，如人員異動、或是職務變更等等情形，則需修改使用者、角色與權限關係，這也大大的增加維護的成本和風險。因此數位典藏權限管控主要的功能，在確保數位典藏環境下，能針對不同的使用者提供合宜的系統服務，以期望能在適當的時間(right time)提供適當的資訊(right information)給適當的人(right people)。為了達到上述的目的，採用使用者認證技術與RBAC權限管控方法，而設計出適合數位典藏環境的權限管控系統架構，如圖3-14所示（張益嘉、林金龍、何建明，民92）。

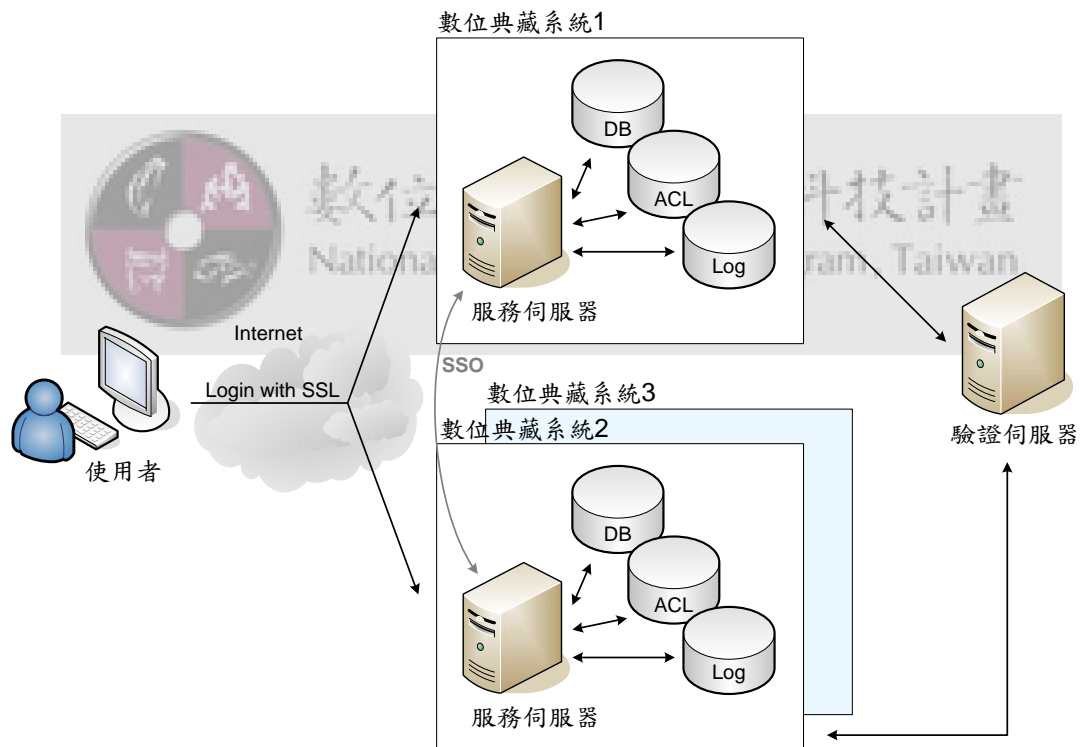


圖 3-14：數位典藏權限管控架構圖

當使用者透過Internet登入任一個數位典藏系統時，數位典藏系統會將使用者驗證資訊傳送至驗證伺服器以進行使用者身份確認。若使用者通過身份認證便可登入典藏系統以進行系統操作。而所有的連線動作都會經過SSL(Secure Sockets Layer)加密傳輸，以確保資料不被竊取。

即使登入成功後，在每次使用者進行系統操作時，都必須再次檢查使用者驗證資訊，若成功才執行使用者的動作；若失敗則輸出錯誤訊息，並取消使用者所執行的動作。故進行系統操作時，可能會發生下列三種狀況：

- (一) 當使用者閒置過久，而使得驗證資訊失效，將顯示錯誤訊息，並重新登入。
- (二) 若使用者所操作的功能未被核准，則顯示使用者並無此項操作的權限。
- (三) 使用者成功的執行系統操作，並回應執行結果。

另外，為了避免使用者在多個數位典藏系統多次登入的費時，架構中也採用重導模式的單一登入機制，讓使用者只要登入一次便可在不同的系統進行切換。數位典藏系統透過共同的驗證伺服器以達到跨系統權限管控功能。由圖 3-14 的數位典藏權限管控架構圖中，可得知每個數位典藏系統是由服務伺服器、資料庫(DB)、權限管控表 (Access Control List, 簡稱 ACL) 與系統日誌(Log)所組成，以下分述之：

- (一) **服務伺服器**：提供數位典藏操作的系統平臺。
- (二) **資料庫**：提供典藏資料儲存的資料庫。
- (三) **權限管控表**：當採用RBAC的權限管控方法時，利用權限管控表來定義系統功能與角色間的對應關係。在網站的環境下，一個完整的系統功能通常需要數個網頁操作流程，所以同一功能下的網頁將會被隸屬於同一個權限。
- (四) **系統日誌**：紀錄使用者在典藏系統中所有的操作行為，以及因操作而影響到的典藏資料異動內容。透過系統日誌將可追查任何時間、任何使用者，所做的任何動作，以提供典藏系統稽核機制之參考。一般系統日誌會包含如下表3-4的內容：

表 3-4：系統日誌內容

	欄位名稱	範例
ID	自動編號	001
RECORD_TYPE	記錄類型	器物
RECORD_NO	記錄識別碼	A0001
LOG_DATE	日期	2006/01/01
PROCESS	執行程序	新增
USER_NAME	使用者帳號	administrator
GROUP_NAME	群組名稱	admin
LOG_CONTENT	異動內容	毛公鼎
REMARK	備註	

以上所提出之權限管控架構，是參考數位典藏環境下對於使用權限的需求而設計，在架構之中也結合現今廣被採用的 RBAC 與 SSO 等概念，因此能夠符合大部份典藏系統的需求。由於不同的典藏單位對於網路安全與使用者認證的政策會有所不同，可以依據上述架構進行修正調整，以符合需求。另外，也可考慮採用 PKI(Public Key Infrastructure) 或 W3C 所制定的 Web Services Security 標準，使本架構更加完備。



第六節 漢字編碼處理

由於資訊技術的演進，促使數位化的工作不再如往常困難，許多機構正嘗試將自有之珍貴的研究資料進行數位化，一則是為了讓這些資料流通得更廣，並提供給更多相關的學者及研究人員使用；二來則是為了使這些資料能多一種保存方式，且具有一樣的內容價值。但是，在數位化過程中，常常會遇到無法輸入之字形，這是因為既有的系統中，漢字交換碼不足，而造成電腦無法處理的問題，這就是所謂的「缺字」問題。此問題在部份漢學古籍或佛學經書當中尤其嚴重，同時也造成上述資料在數位化流通和保存上，有著極大的阻礙（莊德明、謝清俊、林晰，民 87）。

為了解決缺字問題，常用的方法是利用中文系統的使用者造字區，選定一個交換編碼，自行新增所缺之字形，以補足缺乏的漢字。但此法並未真正解決缺字問題，並衍生更多相關問題，例如：造字的管理並不容易、使用者造字區內沒有足夠的空間再容納缺字、使用者造字無法多人共用等。所以為了真正解決缺字問題，及避免上述使用者造字方式所帶來的種種缺點，目前，坊間已發展出幾個相對應的解決方案，如行政院主計處電子處理資料中心建置的「CNS11643 中文標準交換碼全字庫」（簡稱全字庫），及中央研究院資訊科學研究所文獻處理實驗室建置的「漢字構型資料庫」。本章節將針對前述「全字庫」及「漢字構型資料庫」作一基本的介紹，讓讀者在面對缺字問題時，也能夠有充分了解，以選擇出適合的解決方案。

一、常用的漢字編碼

由於各國文字在字數、字型、使用方式和文化上存在差異，因此自從人類開始用電腦處理資訊以來，就面臨了不同國家間資料交換的困擾。尤其臺灣、大陸、南韓與日本所用的都是漢字，就編碼的理論而言，會將每一個字以一個編碼來表示。目前國際間有許多常見的漢字編碼，其中支援中文部份的有 BIG-5、GB、ISO 10646 及 UNICODE，其特色及比較如下表 3-5：

表 3-5：常見的中文編碼標準比較表

	BIG-5	GB	ISO 10646	Unicode
特色	只包含繁體中文字	內碼編排與 ISO 10646 國際編碼標準不同	統一東亞國家所訂立的漢字標準，包含所有 BIG-5 及 GB 的	與 ISO 10646 相同

			字	
漢字字符數目	約 13,000 個	6,763 個(GB-2312)	約 27,000 個	約 27,000 個 (Unicode3.0)
普及地區	臺灣、香港	中國大陸	世界各地	世界各地
開發者	臺灣五大中文軟體公司	中國大陸政府	國際標準化組織 (ISO)	Unicode Consortium

(一) BIG-5 碼：

由資策會於 1984 年與五間有意願共同推動電腦中文化的資訊公司所共同創立的一套編碼標準，故稱大五碼。此編碼標準收納了 13,051 個繁體中文字及 408 個符號，其中包含所謂的常用字 5,401 個，次常用字 7,652 個。在 2003 年財團法人中文數位化技術推廣基金會接受經濟部標準檢驗局委託，召集國內業者代表、專家和學者，就 BIG-5 編碼字元予以重新編制，並正式收錄成為國家標準。除了臺灣以外，其他使用繁體漢字的地區，如香港、澳門，都普遍使用 BIG-5 碼。這已經成為正體中文編碼的標準格式。

(二) GB 碼：

GB 碼是中國大陸所制定的漢字編碼標準。最常使用的電腦編碼標準為 GB2312-80，簡稱 GB-2312。其中收錄了 6,763 個漢字，但未能涵蓋繁體中文字、部分人名、方言、古漢語等的罕用字。因此 GB 是一個簡字的漢字編碼，目前通行於中國大陸地區；新加坡等地也採用此編碼。

(三) ISO 10646：

ISO 10646 是由國際標準組織制訂的編碼標準，其目的是想集結全球通用的字符集，成為一個大聯集，以滿足各國資訊交換的需求。因此 ISO 10646 包含世界上主要語言的字符。其中的漢字部分，是由臺灣、中國、日本和韓國所訂立的漢字編碼標準，共收納了約有 27,000 個漢字。

(四) UNICODE：

Unicode 是 Universal Code 的縮寫，是由一個名為 Unicode 學術學會的機構制訂的字符編碼系統，與 ISO 10646 的概念相同。源起於 1991 年，Unicode 學術學會與 ISO 國際標準化組織共同制訂一套適用於多種語言文本的通用編碼標準。自此以後，兩個組織便保持緊密合作，共同提供全球語言文字與符號之表示、傳送、交換、處理、儲存、輸入和顯示的共同編碼標準，不但避免了資源的浪費，並且真正落實了統一全球文字交換標準的理想。

Unicode3.0 版本在 2000 年推出，當中收納了 49,194 個來自世界各地不同語

言的字符，其中包含 27,484 個漢字字符。在隔年又發表了 Unicode3.1 版，除了 Unicode 3.0 版中的 49,194 個字符，又增加了 44,946 個新字符，其中 42,711 個為漢字字符。

二、全字庫

民國 69 年 9 月，行政院國家科學委員會集合國內編碼專家、學者，達成初步原則，並據此報請行政院核定國家中文資訊標準交換碼編碼原則，經過多次會商，在民國 75 年 8 月正式由中央標準局審定頒布為國家標準，編號「CNS11643」，取名為「中文標準交換碼 (Chinese Standard Interchange Code)」。最後在民國 87 年，由行政院主計處電子處理資料中心負責建置「CNS11643 中文標準交換碼全字庫」網站（簡稱全字庫），如下圖 3-15，其主要的功能如下：



圖 3-15：全字庫網站

- (一) 解決個人電腦中文字數不足的問題：個人電腦(PC)上常用的內碼如 BIG-5、Unicode 碼，其基本中文字集（13,053 或 2 萬餘字）往往不敷使用。全字庫建有字型下載機制，在遇電腦有缺字時，可立即經由網際網路下載字型，取代人工造字方式，減省使用者造字時間。
- (二) 解決自造字交換的問題：以電腦傳送的文件如電子郵件、公文等如使用到自造字時，因編碼不一而無法正確呈現。CNS11643 國家中文標準交換碼收編

字數達七萬餘字，字數充足，故以 CNS11643 傳輸，可以解決自造字交換的問題。全字庫網站上也提供「轉碼工具」，可供使用者進行將 CNS 與常用中文碼之互轉。

- (三) **解決機關、企業、團體內部同字不同碼的問題**：使用者各自造字或下載字型後，造成機關、企業內部自造字「同字不同碼」現象，平添繁複之轉碼工作。全字庫網站上建有「造字分享」機制，可供機關或企業內部所有個人電腦安裝相同的自造字，以維護「同字同碼」原則。
- (四) **協助機關、企業、團體整合及管理個人電腦上的中文字集**：為持續維護「同字同碼」原則，機關、企業或團體內必須實施自造字專人轉管作業，原有舊自造字及後續產生的新增字，均由自造字管理人員負責整合，統一編碼、下載或造字。全字庫網站上建有各項管理工具，可供快速、有效整合及管理內部所有自造字。
- (五) **解決網頁上自造字顯示的問題**：網頁上用到個人自造字時，瀏覽者將無法看到或正確的呈現。全字庫網站上建有「字型即時顯示」機制，可供使用者以適合的字體、顏色、大小顯示在螢幕上。

全字庫目前可供查詢的字共 76,067 個中文字及 894 個符號。76,067 個中文字，可經由網際網路以總筆畫數、注音符號、倉頡碼、拼音及複合等方式查詢其注音、部首、筆畫、CNS11643 及 BIG-5 (含 BIG-5E)、Unicode 等中文碼，並提供 40x40 明體點矩陣字型及楷體向量字型下載。另外，針對檔案交換的需求，全字庫也提供了轉碼工具，可進行純文字檔中常用內碼 BIG-5、EUC、Unicode、GBK (繁體字) 與 CNS11643 的互轉，以達成資訊的正確交換。

三、漢字構型資料庫

對於缺字問題，中央研究院資訊科學研究所文獻處理實驗室發展出漢字構形資料庫，目前收錄了大約六萬個楷體字，並且包含小篆、金文、甲骨文、楚系文字等古文字。這個資料庫除了銜接古今文字外，最重要的是記錄了每個字的字形結構。由於每個字形的結構皆不相同，用來表達字形結構的構字式正好可用來當作字形的識別碼。構字式是由有限的構字部件及符號組成，較目前一字一碼的交換碼更適合用來解決漢字的缺字問題。透過構字式與漢字構形資料庫的運用，中文字在現有電腦環境下不再有數量限制，大量中文字已經可以自然輸入與呈現，可應用在網頁處理或 Microsoft Office 套裝軟體中使用。其詳細的內容如下：

(一) 部件

當一個形體用來構成某字形的一部分時，稱該形體為某字的部件。如：「日」、「京」是「景」的部件；「景」、「頁」是「顛」的部件；「顛」是「灑」的部件；又如：「圍」的部件有「口」和「韋」。因此部件是漢字的構形單位。部件是有層次的，如：「顛」可拆分成「景」與「頁」，「景」又可拆分成「日」、「京」。部件裡有的是字，有的不是字（是不是字，是相對於某字集而說的）。最後不再拆分的最小形體叫作字根。

漢字最常用的構字方法為橫連（ Δ ）、直連（ \oplus ）與包含（ \triangle ）。如將上例寫作：『灑= $\dot{\Delta}$ 顛』，『顛=景 Δ 頁』，『景=日 \oplus 京』，『圍=口 \triangle 韋』，則這些式子稱為「構字式」。一個構字式裡只能有一個拆分的符號，故對缺字而言，構字式表達的是一個唯一的字形，不會與其他文字混淆；因此，以構字式作為缺字的識別碼（identifier）十分理想。

(二) 部件序

由於某些字無法用構字式表示，因為構字式限定只能用一個運算符號。如「牖」字，其的左邊為「片」，右邊的「戶 \triangle 甫」並不在五大碼中，字形組合的方式是『牖=片 Δ （戶 \triangle 甫）』，其中同時用到兩個運算符號，所以不符合組字規則。另一個例子是「爽」字，由「大」、「百」、「百」三個部件組成，但是卻無法採用橫連、直連或包含的方式來分解。故當某些字缺乏適當的部件，如：「牖」字；或缺乏適當的運算符號時，如：「爽」字，則可完全摒除運算符號，用部件序來表示，即直接按照書寫順序的部件表示，並在前後加上起始「 \square 」及終結「 \square 」標示（Tagging）。例如：『牖= \square 片戶甫 \square 』，其中「 \square 」及「 \square 」也是自創的符號。

(三) 方便符號與從缺符號的運用

並不是所有的缺字，都可以利用上述的規則繕打出來，這些缺字中有可能出現沒有提供任何一個部件，或是不知道如何輸入的情形，所以再造一個符號「 $\textcircled{?}$ 」，表示從缺，例如：『竊= \square 穴采 $\textcircled{?}$ \square 』。目前，一個組字式中，最多只允許出現一次從缺符號。若是使用倉頡輸入法的人，或許會覺得「 $\textcircled{?}$ 」與倉頡中的「難」有些類似。

為了方便輸入，另外還造了一些方便符號。方便符號需放在部件的前頭，如：「 ∞ 」表示兩個相同的部件橫連（例如：『競= ∞ 克』）；「 $\textcircled{8}$ 」表兩個相同的部件直連（例如：『爻= $\textcircled{8}$ 戈』）；「 ∞ 」表三個相同的部件橫連，（例如：「 ∞ 去」，中文大辭典編號 3171）；「 $\textcircled{8}$ 」表三個相同的部件直連，（例如：「 $\textcircled{8}$ 戶」，中文大辭典編號 12029）；「 $\textcircled{8}$ 」表三個相同的部件三角頂立（例如：『轟= $\textcircled{8}$ 車』）；「 ∞ 」表四個相同的部件橫連；「 $\textcircled{8}$ 」表四個相同的部件直連；和「 $\textcircled{8}$ 」表四個相同的部

件分佔四方（例如：『燧=ㄨㄨ火』）等。

從缺符號和方便符號是可以用於構字式，如此可擴大構字式的適用範圍；例如，原先只能用部件序表達的字即能改為構字式，如：『瞿=〇〇目厶隹』、『俎=ㄨ人厶且』、『桑=ㄨ又厶木』、『啜=口厶ㄨ又』等。方便符號也可用於部件序中，讓組字更為方便，例如：『獸=ㄨㄨ又酉欠〇』。方便符號也讓一些字的組字式既不需橫連、直連及包含符號，也不需起始及終結標記，例如：「ㄨ魚」（中文大辭典編號 47603）。

（四）缺字序號

有些字連一個部件都拆不出來，以中文大辭典 49,905 個字為例，其中約二千字很像圖形，幾乎無法拆分，這時候就得利用缺字序號來作識別。缺字序號的型式如同部件序，需要用〈起始〉和〈終結〉的標示，但是標示中的部件改為編號，如：『ㄨ 5 〇』表示這是第五個無法利用前規則表達的缺字。



參考文獻

- 林彥君、黃建中、王祥安(民94)。數位典藏軟體需求規格書建議標準。**臺灣軟體工程研討會**，第一屆，323-329。
- 范紀文(民91)。數位典藏技術整合與運用。**數位典藏技術說明會及成果展**。上網日期：民95年12月10日。網址：
<http://datf.iis.sinica.edu.tw/Meeting/030602-fann.pdf>
- 范紀文、何建明、李德財(民90)。從典藏資料交換角度探討 Metadata 設計與標準化問題。**新世紀數位圖書館與數位博物館趨勢研討會**。新竹市：交通大學圖書館。上網日期：民95年12月10日。網址：
<http://daal.iis.sinica.edu.tw/pdf/3-2.pdf>
- 范紀文、何建明、李德財(民92)。數位典藏管理系統分析與設計-以故宮書畫典藏為例。**數位典藏技術研討會**。上網日期：民95年12月10日。網址：
<http://daal.iis.sinica.edu.tw/pdf/A-3.pdf>
- 張益嘉、林金龍、何建明(民92)。數位典藏系統之權限控管。**數位典藏技術研討會**。上網日期：民95年12月10日。網址：<http://daal.iis.sinica.edu.tw/pdf/E-4.pdf>
- 莊德明、謝清俊、林晰(民87)。中央研究院古籍全文資料庫解決缺字問題的方法。**第二次兩岸古籍整理研究學術研討會**。上網日期：民95年12月10日。網址：http://www.sinica.edu.tw/~cdp/paper/1998/19990511_1.htm
- 陳亞寧、陳淑君、沈漢聰、鍾豐謙(民91)。後設資料系統的需求評選與發展設計。**數位典藏技術研討會**。上網日期：民95年12月10日。網址：
http://www.sinica.edu.tw/~metadata/bibliography/proceeding/paper/ndap_dats2002pp.pdf
- 黃國倫、李家豪、何建明(民92)。漢代簡牘數位典藏系統之設計與實作。**數位典藏技術研討會**。上網日期：民95年12月10日。網址：
<http://daal.iis.sinica.edu.tw/pdf/B-2.pdf>
- 黃國倫、蕭人豪、李家豪、陳心渝(民93)。數位典藏網頁缺字解決方案及應用。**數位典藏技術研討會**，第三屆。
- 黃景彰(民89)。**使用XML設計執行權管制資訊流**。未出版碩士論文，國立交通大學資訊管理研究所。
- 廖英彥(民94)。**網際網路單一簽入系統應用**。未出版碩士論文，世新大學資訊管理學系。
- 廖鴻圖、林金龍(民94)。以P2P網路架構實作Single Sign-On機制之研究。**數位典藏技術研討會**。上網日期：民95年12月10日。網址：
<http://daal.iis.sinica.edu.tw/pdf/2005C-2.pdf>
- 漢字構形資料庫 <http://www.sinica.edu.tw/~cdp/>
- 鍾子帆、何建明(民91)。數位典藏系統之多媒體檔案管理與呈現。**數位典藏技術研討會**。上網日期：民95年12月10日。網址：

<http://daal.iis.sinica.edu.tw/pdf/F-2.pdf>

Chen, Y.-N., Chen, S.-J., & S.C. Lin. (2003). **A metadata lifecycle model for digital libraries: Methodology and application**. Paper presented at World Library and Information Congress: 69th IFLA General Conference and Council, 1-9 Aug. 2003, IFLA. Berlin, Germany: ICC. Available at:
http://www.ifla.org/IV/ifla69/papers/141e-Chen_Chen_Lin.pdf

關鍵字列表

HTTP 重導(HTTP Redirect)

Java 程式語言整合(Java Binding)

PKI(Public Key Infrastructure)

SSL(Secure Sockets Layer)

Unicode (Universal Code)

W3C

Web Services Security

中文標準交換碼(Chinese Standard Interchange Code)

主從式架構(Client Server)

主鍵(Primary Key)

以個體為基礎(Identity-based)

以規則為基礎(Rule-based)

以網頁為基礎(Web-based)

代理人模式(Proxy Model)

外來鍵(Foreign Key)

用戶端(Client)

目標使用者(Target Users)

存取權限(Access Right)

存活機制暨評估 (Survivability Assessment)

系統日誌(Log)

角色為基礎的使用權管控((Role-based Access Control , 簡稱RBAC))

防火牆 (Firewall)

使用者(User)

使用權管控(Mandatory Access Control , 簡稱MAC)

服務伺服器(Service Server)

服務請求(Service Request)

阻絕癱瘓攻擊 (Denial of Service Attack , 簡稱DoS)

客戶(Customer)

重導模式(Inform Model)

個人電腦(PC)

原生型XML資料庫 (Native XML Database)

授權及權限管理架構 (Privilege Management Infrastructure)

軟體需求規格書(Software Requirements Specification, 簡稱SRS)

單一登入(Single Sign-On)

登入主機 (Login Server)

損害範圍隔離策略(Defect Isolation)

會議金鑰(Session Key)

資料庫(DB)

網站(Web Site)

網站主機(Web Server)

網站系統伺服器(Web/Application Server)

網際網路(Internet)

標示 (Tagging)

標籤(Tag)

請求(Request)

適當的人(Right People)

適當的時間(Right Time)

適當的資訊(Right Information)

學習物件(Learning Object)

隨意型的使用權管控(Discretionary Access Control , 簡稱DAC)

識別碼(Identifier)

權限管控表(Access Control List , 簡稱ACL)

驗證主機(Verification Server)

驗證伺服器(Authentication Server)