

從搜集、整理、建置- 談圖書館的數位典藏生命週期

林彥宏

中央研究院歷史語言研究所

數位典藏與數位學習國家型科技計畫

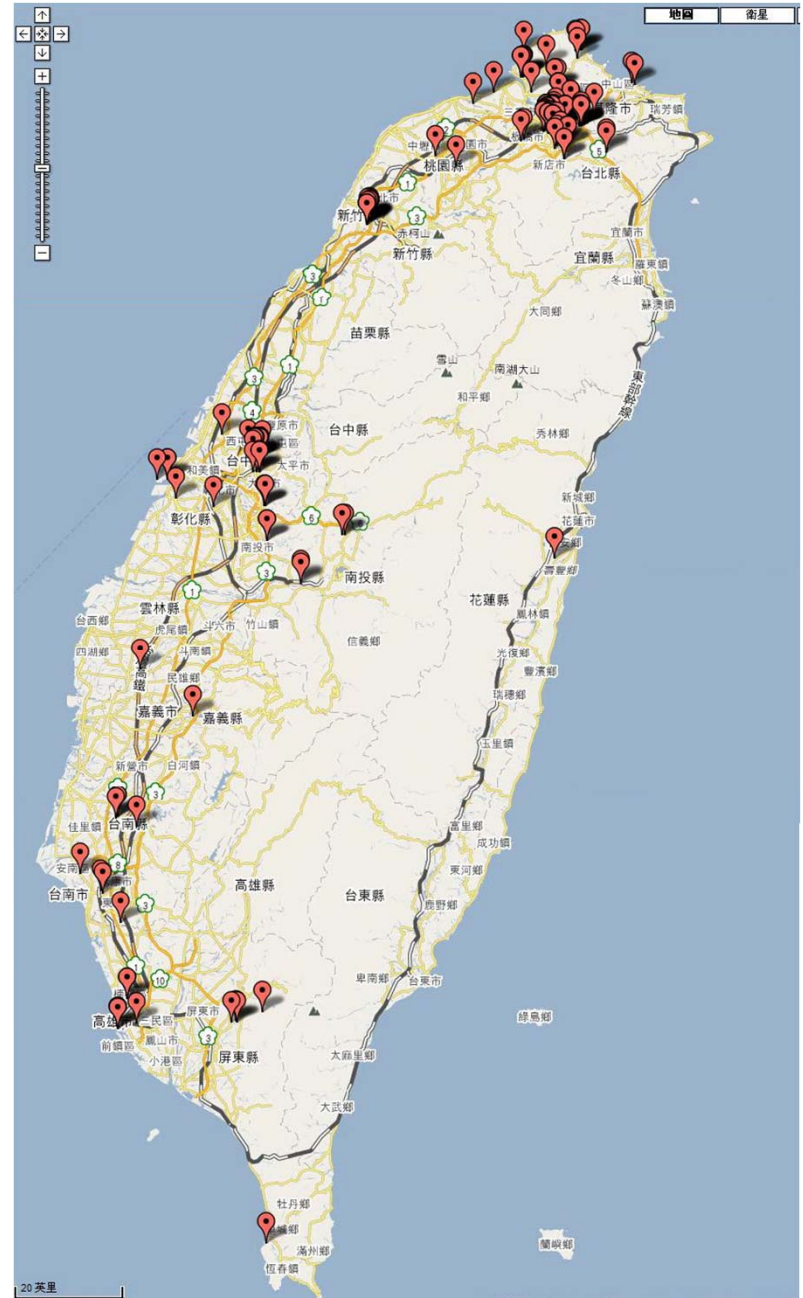
拓展台灣數位典藏計畫

執行秘書

yenny@gate.sinica.edu.tw

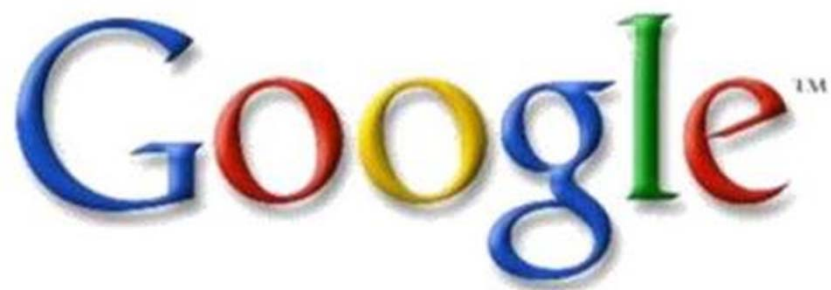
數位典藏 內容建置 單位

中央研究院
科學博物館
故宮博物院
臺灣大學
國史館
臺灣文獻館
國家圖書館
國家電影資料館
檔案管理局
原住民族委員會
客家委員會
大專院校
公私立典藏單位



<http://content.teldap.tw/>

數位時代 資訊搜尋



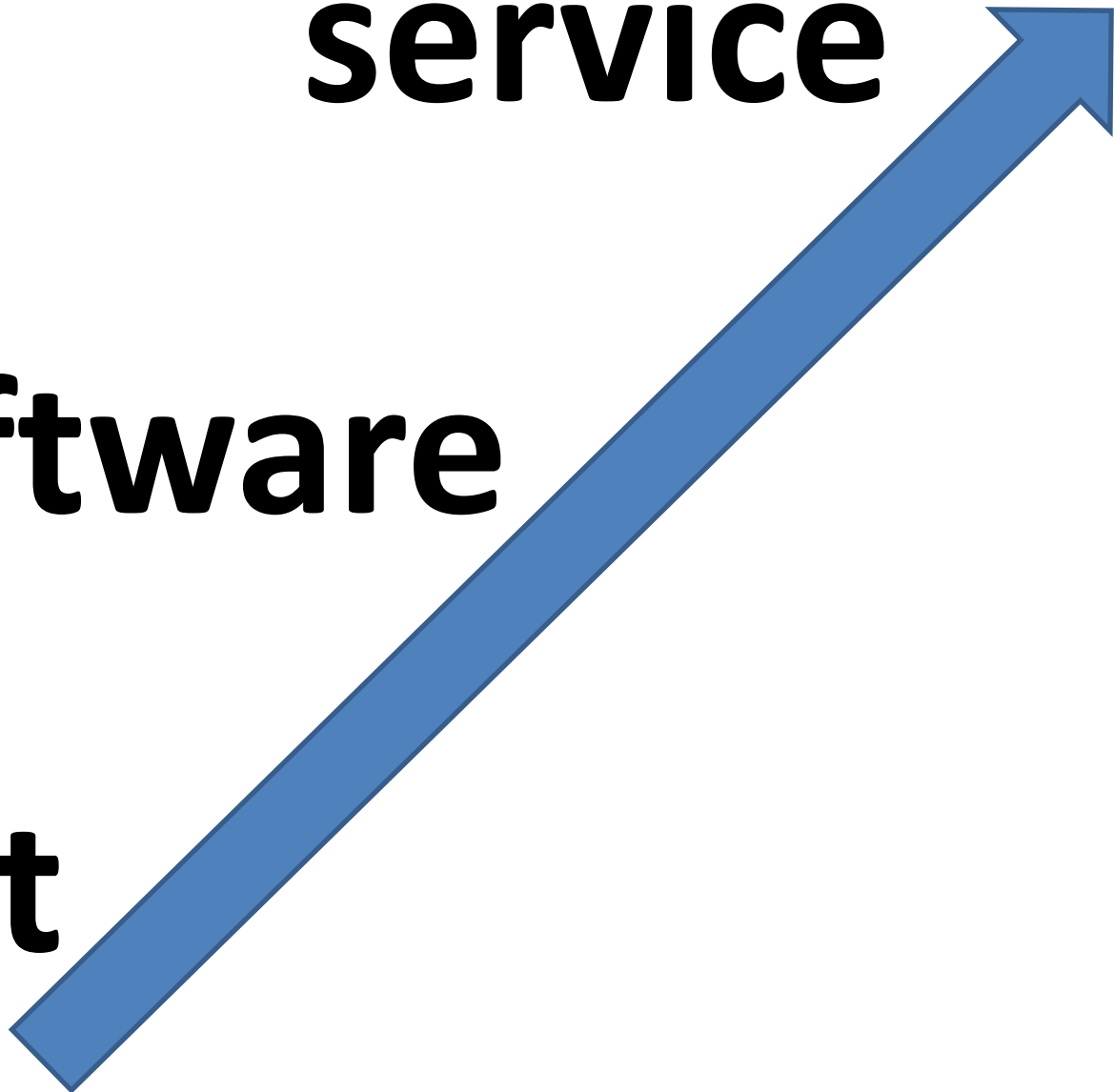
facebook

從資訊組
織到數位
典藏

service

software

content



數位典藏 生命週期



藏品數位化

- 藏品精選
- 數位化規格
- 數位化轉換

制訂藏品精選政策

篩選標準	考慮問題
智慧財產權	<ul style="list-style-type: none">• 是否具有藏品或作品的著作權• 是否具有數位化和傳播的法定權利• 是否能夠控制數位化版本的使用
需求分析	<ul style="list-style-type: none">• 是否具有足夠的軟、硬體• 是否具有適當的技術• 是否具有足夠的人力資源
執行預算	<ul style="list-style-type: none">• 機構/單位是否具有預算• 是否需要外部經費補助• 藏品是否可吸引外部資金

篩選標準	考慮問題
藏品價值	<ul style="list-style-type: none">• 是否具獨特性• 是否有歷史意義• 是否具代表性• 是否具有完整性• 是否對某主題領域的發展有重要性
藏品維護	<ul style="list-style-type: none">• 是否具有安全性的顧慮• 是否有效減少原件的調件數• 提升原件的可讀性
使用者需求	<ul style="list-style-type: none">• 誰是預期的使用者• 數位化是否可吸引新的使用者• 未來是否具有潛在使用者或合作伙伴
對圖書館的影響	<ul style="list-style-type: none">• 對圖書館的運作是否具有重要性• 與機構發展的策略是否相符

數位化規格舉例

國家圖書館			
實體物件狀況	若文字線條過細、文字太小、其他特殊要求	古籍原書 B4-A3尺寸	普通線裝書 B4-A3尺寸
比例	1 : 1		
影幅取捨	封面獨立一頁，其餘左右半頁合為一頁		
檔案格式	TIFF	TIFF	TIFF
色彩模式	RGB (24bit/pixel)	RGB (24bit/pixel)	黑白
解析度及尺寸	原尺寸600dpi以上	原尺寸300dpi以上	原尺寸300dpi以上

解析度代表數位影像所內含的像素數量，與構成影像的水平 and 垂直的像素數量有關。通常採用每英寸的距離內有多少像素來衡量，簡稱ppi (Pixel Per Inch，每英寸像素數)



30 ppi



60 ppi



150 ppi



相機類型	數碼單鏡反光相機
影像感應器	藝康 DX 格式 (23.6 x 15.8 毫米) CMOS 感光元件
有效像素	12.3 百萬像素
影像大小 (像素)	4,288 x 2,848 [L], 3,216 x 2,136 [M], 2,144 x 1,424 [S]

產品類型	平台式彩色掃描器
影像感應元件	線性 CCD 與冷陰極管
掃描方式	單次單面掃描
解析度	光學解析度：600 dpi
掃描速度	彩色模式 200 dpi 解析條件下，每頁掃描(A3 全幅) 3 秒以內

pixel – 像素(輸入端)

dot – 影像質點(輸出端)

ppi - 虛擬/掃描解析度

dpi - 實體解析度

lpi - 網線數

範例：

5×3英吋的老照片，要
輸出10×6英吋大小，
需要的解析度為多少？

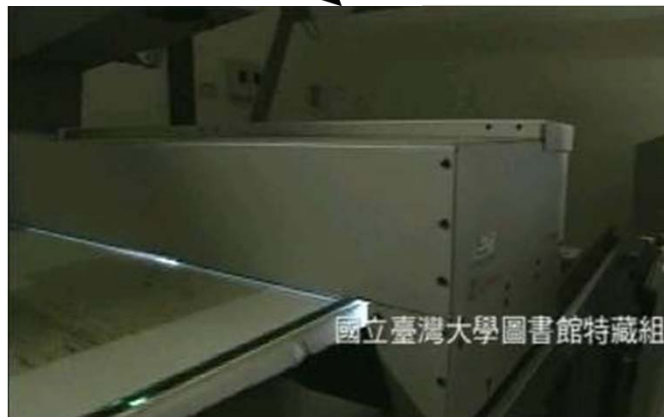


解析度 = 像素數量 ÷ 尺寸

輸出端 $(10 \times 300\text{dpi}) \times (6 \times 300\text{dpi}) = 3000 \times 1800$
= 540萬像素
= 擷取端 $(5 \times 600\text{ppi}) \times (3 \times 600\text{ppi})$

主檔影像	使用影像	縮圖影像
<ol style="list-style-type: none"> 1.盡可能描繪原件的資訊內容 2.無壓縮 3.未編輯 4.作為衍生檔案的長期來源 5.作為原件的代用品 6.高品質 7.非常大的檔案尺寸 8.用來創建高品質的印刷複製 9.通常以TIFF檔案格式儲存 	<ol style="list-style-type: none"> 1.代替主檔影像作為一般網路使用 2.通常符合一般螢幕觀看 3.合理的檔案尺寸，可加速下載時間 4.對一般研究而言是可接受的品質 5.壓縮以加速存取的速度 6.通常以JPEG檔案格式儲存 	<ol style="list-style-type: none"> 1.非常小的影像，通常作為目錄紀錄 2.在線上可快速展示，並允許使用者連結至「使用影像」 3.通常以GIF 或JPEG 檔案格式儲存 4.不適合文字或樂譜為主的影像，在如此小的等級，使用者無法辨別內容

數位影像 製作



國立臺灣大學圖書館特藏組

品質檢驗與管理

1. 影像尺寸的優化 – 如裁切、方向校正等
2. 影像色調的優化 – 如白平衡、色階和曲線等
3. 清潔 - 除去記號、污點、改正影像的其他問題
4. 增加或編輯影像固有的IPTC後設資料，包括版權標籤
5. 從RAW或高位元TIFF變換到24bit RGB，TIFF無壓縮檔案格式
6. 如果有必要，可以使用一些有限的銳利化，但應該記錄在技術後設資料內

數位化工作流程指南

97年出版指南



文書檔案



地圖圖資



期刊報紙



書畫



古籍線裝書



瓷器



照片



色彩管理



委外製作

98年出版指南



考古資料



正片



水下生態影片



維管束植物



礦物平面攝影



語料庫建置入門



印章印記



專案規劃



整合性工作流程

- 參考網站 <http://content.teldap.tw>

建置後設資料

- 何謂後設資料
- 後設資料品質評估

後設資料 舉例

欄位	內容
標題	文選
作者	(南北朝)蕭統(編)
類型	善本古籍
發行者	建陽崇化書坊陳八郎宅
發行日期	宋紹興辛巳三十一年(1161)
語言	中文
權利	國家圖書館
描述	《文選》又稱《昭明文選》，為南朝梁武帝長子蕭統主編，為現存最早的中國文學總集。此為宋刻單行五臣注《文選》，較尤袤刻李善注早十九年。自六臣注合本出，而五臣單注本失傳，四庫提要亦稱未見此本，實屬罕見秘籍。書內朱筆亦出宋人手，凡宋諱嫌名，如懸、炯……等字，皆有紅圈，當為寧宗時人讀本。宋刻宋印，而又宋讀，真人間奇書。

資料結構

data structure

欄位	內容
標題	文選
作者	(南北朝)蕭統(編)
類型	善本古籍
發行者	建陽崇化書坊陳八郎宅
發行日期	宋紹興辛巳三十一年(1161)
語言	中文
權利	國家圖書館
描述	《文選》又稱《昭明文選》，為南朝梁武帝長子蕭統主編，為現存最早的中國文學總集。此為宋刻單行五臣注《文選》，較尤袤刻李善注早十九年。自六臣注合本出，而五臣單注本失傳，四庫提要亦稱未見此本，實屬罕見秘籍。書內朱筆亦出宋人手，凡宋諱嫌名，如懸、炯.....等字，皆有紅圈，當為寧宗時人讀本。宋刻宋印，而又宋讀，真人間奇書。

標準	簡介	範例
Categories for the Description of Works of Art (CDWA/CDWA-Lite)	<p>由蓋提基金會藝術維護資訊任務小組維護，提供博物館與檔案學界描述藝術作品（如繪畫、雕刻、陶藝、建築、傢俱、表演藝術等）及其影像資料的標準。CDWA所定義的架構(共有532項類目)也提供交換和共享的功能，使各地的典藏機構可以交換藏品資訊。CDWA-Lite為基於資料交換目的而提出的XML schema，包含22項核心類目。</p>	<ul style="list-style-type: none"> • 國立故宮博物院 • 台灣多樣性知識網
Guidelines for Electronic Text Encoding and Interchange (TEI)	<p>TEI是一國際性與跨學科性的標準，協助圖書館、博物館、出版者和個別學者以電子文本形式呈現各種文獻和(語言學)文本，以達到線上教學與利用。</p>	<ul style="list-style-type: none"> • 中華電子佛典協會
Dublin Core Metadata Element Set (Dublin Core / DC)	<p>DC的提倡，是為了加速網路電子資源的整理與組織，並加強網路資源的找尋與檢索的精確性，其15項元素具有核心和通用的特性。</p>	<ul style="list-style-type: none"> • 數位典藏聯合目錄

資料內容

data content

欄位	內容
標題	文選
作者	(南北朝)蕭統(編)
類型	善本古籍
發行者	建陽崇化書坊陳八郎宅
發行日期	宋紹興辛巳三十一年(1161)
語言	中文
權利	國家圖書館
描述	《文選》又稱《昭明文選》，為南朝梁武帝長子蕭統主編，為現存最早的中國文學總集。此為宋刻單行五臣注《文選》，較尤袤刻李善注早十九年。自六臣注合本出，而五臣單注本失傳，四庫提要亦稱未見此本，實屬罕見秘籍。書內朱筆亦出宋人手，凡宋諱嫌名，如懸、炯.....等字，皆有紅圈，當為寧宗時人讀本。宋刻宋印，而又宋讀，真人間奇書。

標準	簡介
中國編目規則	參考英美編目規則之目錄著錄原則，並考慮到我國目錄著錄之實際需要而研訂。
Resource Description and Access (RDA)	RDA 是以IFLA發展的兩個觀念模型——《書目記錄的功能需求》(FRBR)和《規範資料的功能需求》(FRAD)為基礎，提供關於數位資源編目的指南，並更強調說明使用者查找、標識、選擇和獲得他們所需要的資訊。RDA 還支援書目記錄的聚類，顯示作品及其創建者之間的關係。這個重要的新特點使得使用者更清楚地瞭解作品的不同版本、翻譯或物質形式。
Cataloging Cultural Objects (CCO)	適用於文物的資料內容標準，以現有的標準為基礎，CCO為編目記錄中的後設資料元素之資料值，提供資料之篩選、排序和格式化的指導。

資料值
data values

欄位	內容
標題	文選
作者	(南北朝)蕭統(編)
類型	善本古籍
發行者	建陽崇化書坊陳八郎宅
發行日期	宋紹興辛巳三十一年(1161)
語言	中文
權利	國家圖書館
描述	《文選》又稱《昭明文選》，為南朝梁武帝長子蕭統主編，為現存最早的中國文學總集。此為宋刻單行五臣注《文選》，較尤袤刻李善注早十九年。自六臣注合本出，而五臣單注本失傳，四庫提要亦稱未見此本，實屬罕見秘籍。書內朱筆亦出宋人手，凡宋諱嫌名，如懸、炯.....等字，皆有紅圈，當為寧宗時人讀本。宋刻宋印，而又宋讀，真人間奇書。

標準	簡介
中文主題詞表	<p>以《中文圖書標題表》為藍本，參考各種關鍵詞庫、專科詞典、主題詞表等，並參酌文獻編目實務經驗，採文獻保證及用戶保證原則，增補了許多反映新學科、新事物、新概念的詞目。</p>
Art & Architecture Thesaurus (AAT)	<p>AAT內含結構性詞彙，用以描述藝術、建築、裝置藝術、物質文化和檔案素材。可作為編目與文獻工作的依據，協助藝術、建築和物質文化方面的資料檢索。此索引典包含約34,000個概念，131,000個詞彙，詞彙與詞彙之間的關係包括等同、層級和聯想三種關係，因此，它是一部符合ISO和NISO標準的索引典。</p>
明清檔案人名權威	<p>人名權威檔係以明清檔案(內閣大庫檔案、軍機檔、宮中檔等)為主要參考資料，另輔以故宮典藏之清國史館傳包、傳稿及清史館傳稿，再選取學界公認具權威性的史料。目前已建置14800餘筆。</p>

後設資料品質評估

- ❑ 符合藏品、使用者、數位物件等方面現今和未來的用途。
- ❑ 後設資料可以支援互通。
- ❑ 使用權威控制和內容標準，提升使用者搜尋結果的準確性。
- ❑ 對於數位化物件的使用條件有明確說明。
- ❑ 能夠支援典藏品的長期管理。
- ❑ 後設資料是正確且應有唯一識別碼。

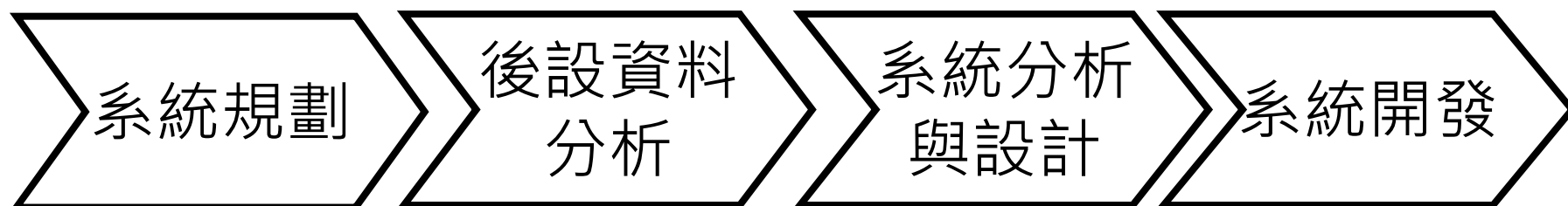
系統展示

- 展示系統舉例
- 系統建置流程
- 以Greenstone為例

展示系統舉例

- 美國記憶 <http://memory.loc.gov/ammem/index.html>
- 美國記憶「學習網」 <http://www.loc.gov/teachers>
- Flickr–The commons <http://www.flickr.com/commons/>
- Internet Archive <http://www.archive.org/>
- 數位典藏聯合目錄 <http://digitalarchives.tw>
- 台灣多樣性知識網 <http://knowledge.teldap.tw>

系統建置流程



典藏單位

後設資料
專家

系統開發人員

系統規格書

後設資料需求規格書

系統需求與設計規格書

數位典藏系統



系統檢核表

其他展示系統類型

- 使用個人或簡單的管理系統

Irfanview

- 使用網路服務的管理系統

Flickr

- 使用商業套裝軟體

eMuseum

- 使用開放原始碼管理系統

Greenstone

長期保存的挑戰

- 資訊科技的發展造成技術的淘汰，每三至五年設備的老舊更換，都是資料長期保存的威脅。
- 儲存媒體容易損壞的特性，脆弱的磁性和光學塗料，皆隨著時間漸損耗。
- 維持數位資料脈絡關係及完整性，方能長久取用正確、完整的數位資訊。
- 維持資料的真實性、可靠性，透過與資料的來源、變動過程結合，保存完整追蹤機制和歷史紀錄。

長期保存的建議

- 盡可能高的品質
- 使用開放標準
- 適當的保存環境定
- 定期檢驗和轉置

報告完畢、謝謝聆聽！！