



# 後設資料對照程序及 品質指標檢測

後設資料工作組(MAAT)

陳亞寧 陳慧婷

7 Sep.,2010



第二部份：對照程序



## 一對一原則

- 目的：釐清描述對象，此為決定後設資料正確性的一大重點
- 描述對象不同，可能出現差異的欄位主要有
  - 創作者
  - 出版者
  - 貢獻者
  - 型式
  - 格式
  - ...等



## 一對一原則

籐編甌



型式

靜態圖像

(描述對象為數位圖檔)

物

(描述對象為原件)



# 一對一原則

格式一 高廣度	畫心尺寸：29×18.8cm (描述對象為原件)
格式一 高廣度	檔案大小：37,251 bytes (描述對象為數位圖檔)

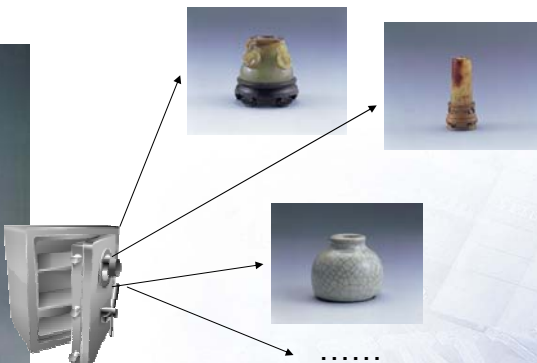
溥心畬與張大千  
合繪山水畫



Source : <http://catalog.digitalarchives.tw/?URN=431089>



# 一對一原則



各式珍玩共47件

Source : <http://catalog.digitalarchives.tw/?URN=836836> ; <http://catalog.digitalarchives.tw/?URN=836837> ;  
<http://catalog.digitalarchives.tw/?URN=836838> ; <http://catalog.digitalarchives.tw/?URN=836842>



# 一對一原則

清 雕紫檀龍紋多寶格



明 青玉四耳活環尊



關聯—有  
一部分為

明 青玉四耳活環尊  
(故玉005759N000000000) ;  
漢 蒼玉王卜勒  
(故玉005760N000000000) ;  
... (略, 共47件)

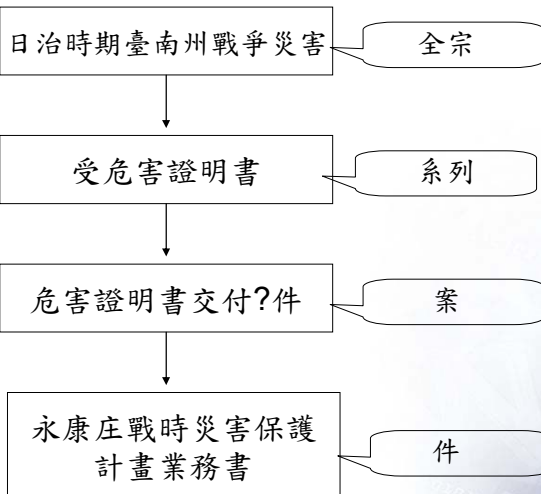
關聯—  
是...的一  
部分

清 雕紫檀龍紋多寶格 (故雜  
001374N000000000)  
(描述對象為多寶格中的其中一  
件玉器, 則指出多寶格的識別  
碼, 表示其為多寶格的一部分)

Source : <http://catalog.digitalarchives.tw/?URN=836836> ;  
<http://catalog.digitalarchives.tw/?URN=836837>



# 一對一原則



Source : <http://catalog.digitalarchives.tw/?URN=456232>



## 一對一原則

永康庄戰時災害保護計畫業務書



關聯一是…的  
一部分

全宗名：日治時期臺南州戰爭災害

系列名：受危害證明書

卷名：危害證明書交付?件

(描述對象為檔案層級中的「件」，則指出其所屬的全宗、系列、卷之名稱)

Source : <http://catalog.digitalarchives.tw/?URN=456232>



## 簡化原則

- 目的：保留原有的資訊
- DC欄位中可使用修飾詞 (qualifier)，或不使用修飾詞
- 建議在匯入資料時，一併匯入計畫原欄位名稱，以資識別資訊。

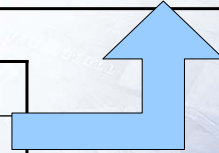


# 簡化原則



標題	文物名稱：婦女頭飾 英文名稱：female head ornament 族語名稱：vanga
----	------------------------------------------------------

文物名稱	婦女頭飾
英文名稱	female head ornament
族語名稱	vanga

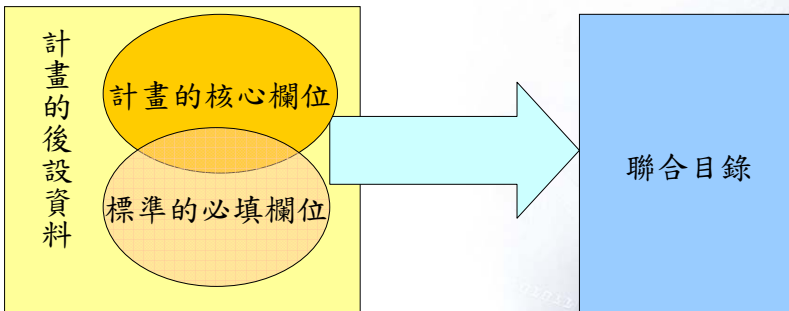


Source : <http://catalog.digitalarchives.tw/?URN=292903>



## 確認計畫的核心欄位或必填欄位

- 目的：保留計畫中重要的核心資訊，以及藏品的故事性
- 確認計畫後設資料的核心欄位或必填欄位，以及若有採用後設資料標準，該標準的必填欄位。這類欄位應作為對照至聯合目錄的首要欄位





## 確認聯合目錄必備欄位

- 目的：達到聯合目錄對於資料品質的最低要求
- 聯合目錄必備欄位包括
  - 識別碼(Identifier)
  - 標題(Title)
  - 主題(Subject)
  - 出版者(Publisher)
  - 格式(Format)
  - 權利(Rights)



## 確認聯合目錄必備欄位

- 若計畫核心欄位或必填欄位不足以滿足聯合目錄必備欄位時
  - 審視其他欄位是否有適合對照至聯合目錄必備欄位者
  - 若原始欄位皆無適合對照至聯合目錄必備欄位時，計畫可增加此欄位資訊，由系統自動或人工匯入





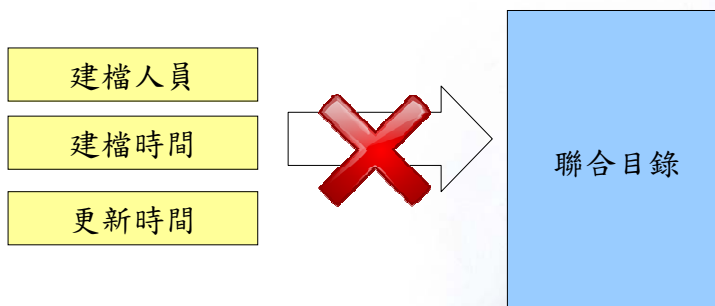
## 優先對照的欄位

- 目的：先進行簡單較不需思考且不易出錯的對照
- 原資料的欄位名稱與定義與DC欄位相符者，優先對照
- 此部分可以簡單對照的主要欄位為
  - 標題
  - 創作者
  - 識別碼
  - 語言
  - ...視各計畫情況而定。
- 對照時除了欄位名稱外，定義需特別注意
  - 有時計畫原始欄位名稱與DC欄位名稱相同，但其定義內涵與使用限制可能會有所差異



## 非所有欄位皆需對照至聯合目錄

- 並非所有計畫的原有欄位皆需對照至聯合目錄，例如內部管理資訊的後設資料以及不適宜開放給一般使用者的資訊等。







## 聯合目錄匯入格式及標點符號使用

- 一併匯入原有欄位名稱，原欄位與資料值之間以全型冒號「：」區隔。

內容摘要	鹿港理番廳同知 陳盛韶移請淡水 廳同知婁雲希將 中港社公務迅飭 劉紀年辦理
------	---------------------------------------------------



描述	內容摘要：鹿港理番廳同知陳盛韶移請淡水廳同知婁雲希將中港社公務迅飭劉紀年辦理
----	----------------------------------------

Source : <http://catalog.digitalarchives.tw/?URN=1901428>



## 聯合目錄匯入格式及標點符號使用

- 若匯入時為多對一欄位時，則以斷行方式區隔

主要主題	拓片
類目	石刻部-祠廟類- 佛寺之屬



主題	主要主題：拓片 類目：石刻部-祠廟類-佛寺之屬
----	----------------------------

Source : <http://catalog.digitalarchives.tw/?URN=1375403>



## 聯合目錄匯入格式及標點符號使用

- 若單一欄位有多值時，則合併並以全形頓號區隔

關係人	買主夥記顧岐山
關係人	為中人簡鼎宗
關係人	知見人男文傑

貢獻者	關係人：買主夥記顧岐山、為中人簡鼎宗、知見人男文傑
-----	---------------------------

Source : <http://dspace.lib.fcu.edu.tw/handle/2377/28363>



## 聯合目錄匯入格式及標點符號使用

- 若欄位有階層式關係時，以全形破折號合併呈現此欄位。

主題分類(第一層)	動物
主題分類(第二層)	脊椎動物
主題分類(第三層)	哺乳類

主題	主題分類：動物－脊椎動物－哺乳類
----	------------------



## 聯合目錄匯入格式及標點符號使用

- 上述格式及標點符號使用原則，可搭配使用

關鍵字	保育
關鍵字	黑猩猩
主題分類(第一層)	動物
主題分類(第二層)	脊椎動物
主題分類(第三層)	哺乳類

主題	關鍵字：保育、黑猩猩 主題分類：動物－脊椎 動物－哺乳類
----	------------------------------------



## 後設資料品質指標的應用與實例

後設資料工作組(MAAT)

陳亞寧 陳慧婷

聯合目錄

溫淳雅



# 大綱

- 說故事的內容重點
- 聯合目錄的資料範例
- 後設資料品質指標的內容簡介
- 資料匯出的必備欄位
- 相關配套措施－對照服務
- 資料對照的範例
- 後設資料品質指標的實例說明（與聯合目錄小組合作）



## 說故事的內容重點

Metadata Elements





# 聯合目錄的資料範例：1974玫瑰花<sup>1</sup>

## 1974 玫瑰花



[後設資料](#) [授權與引用資訊](#) [連結到原始資料網站](#)

**資料類型：**

原件；美術

**著作者：**

李石樵

**主題與關鍵字：**

玫瑰花／靜物／花瓶

寫實靜物

西方媒材／油彩

**描述：**

說明：在回返寫實後，李氏不再刻意追求形的變化，而以自然形象所涵蓋之結構為主，但有時並不明顯顯出經營痕跡。此作雖描繪簡單的靜物，但極寫實的表象卻奠基於幾何結構上，玫瑰花錯落有致，但自然形成一圓圈，此外線圓圈又為葉子所形成之內線所呼應，花瓶之渾圓就不言而喻了。背景之方塊嵌板關係著意為之，因此方塊與畫面左右不相連屬。桌面自然顯示方塊，而其上的方塊圖案則與桌面和背景之方塊呼應，在簡單中見複雜。

材質：畫布／油彩

**出版者：**

藝術家出版社／台北縣立文化中心

(續下頁)



# 聯合目錄的資料範例：1974玫瑰花<sup>2</sup>

## 1974 玫瑰花



**日期：**

創作日期：1974

**格式：**

典藏實體：1件；72.5 × 60.5 cm

**關聯：**

李石樵著，黃明政主編，林宗鼻、千音、黃玉淇攝影，《李石樵畫集》，1996，台北：台北縣立文化中心／李石樵繪；王德育著，《李石樵》，臺灣美術全集（8），1993，臺北市：藝術家出版社

**認圖：**

創作日期：1974

內容地點：台北

**管理權：**

財團法人李石樵美術館基金會



## 聯合目錄的資料範例：魚型燈心袋

中文品名：魚型燈心袋

後設資料 授權與引用資訊 連結到原始資料網站

資料識別：

登錄號：88-00097

資料類型：

編織

格式：

縱：78

橫：22

寬：3

關聯：

朝代：清末民初



## 後設資料品質指標的內容簡介

符合預期  
Expectation/Output

來源  
Provenance

完整  
Completeness

正確  
Accuracy

取用  
Accessibility



## 後設資料的品質指標<sup>1</sup>

指標	定義	算式	範例	權重正規化
O.I.1. 藏品數量	計畫「當年度」 <b>承諾</b> （於管考報告上填寫的「 <b>實際匯入</b> 」數）新增多少筆藏品XML	計畫當年度承諾匯入之藏品XML總數	1200	O類指標為 <b>預期符合性品質指標</b> ，為雷達圖第一部份。以年度新增藏品XML總數與計畫書承諾達成的匯入藏品XML總數相除，求取提交率（O.I.2.）得O部分的分數 範例：106.67%
O.I.2. 藏品匯入比例	計畫「當年度」 <b>確實</b> 新增匯入筆數佔承諾匯入總數之比例	年度 <b>確實</b> 新增藏品XML總數 ÷ 計畫當年度 <b>承諾</b> 匯入之藏品XML總數	1280	



## 後設資料的品質指標<sup>2</sup>

指標	定義	算式	範例	權重正規化
P.I.1. 提供單位資訊	<b>Rights or Publisher</b> 欄位是否填值？	<b>Rights or Publisher</b> 欄位非空白的XML總數 ÷ 藏品XML總數	1280 ÷ 1280 = 100%	P類指標為產生來源性品質指標，為雷達圖第二部份。四項指標（P.I.1, P.I.2, P.II.1, P.II.2）各佔25%，加總得P部分的分數。人工判斷部分，由UC提供50筆資料的連結給MAAT，再由MAAT抽樣進行人工判斷。 範例：81.25%
P.I.2. 元素著錄規範的採用	是否有採用著錄規範進行元素的著錄？如：CCO、AACR/RDA、ISAD (G)。或內部自行發展的著錄規範？	人工判斷	75%	
P.II.1. 控制詞彙、索引典或權威檔的採用	是否有使用控制詞彙、索引典或權威檔等現有標準，以著錄特定元素的內容？如：LCSH、LCNAF、AAT、TGN、ULAN等。或內部自行發展的控制詞彙、索引典或權威檔，乃至於代碼表？	人工判斷	75%	
P.II.2. 記錄元素資料值之著錄規範	是否採用著錄規範進行記錄元素資料值的著錄？如：AACR、DACS、CCO等。或內部自行發展的著錄規範？（L1）	人工判斷	75%	





## 標準規範評分

- A. 採用著錄規範（國際標準或自訂），資料內容品質一致且良好；
- B. 未宣稱採用任何著錄規範，但資料內容品質一致且良好。
- C. 採用著錄規範（國際標準或自訂），但實際資料內容品質有缺失（如缺乏必填欄位等）。若歸納至此分級則評等者必須加入評注，說明詳細的缺失情形及建議的改善方式。
- D. 未宣稱採用任何著錄規範，且資料內容品質不良有缺失。
- 採用標準: A-100 |B-95 |C-75 |D-70



## 後設資料的品質指標<sup>3</sup>

指標	定義	算式	權重正規化
C.I.1. 元素使用完整性	Title = 計畫有使用Title的藏品XML總數 ÷ 計畫藏品XML總數 DC十五個元素都計算一次	Title = 計畫有使用Title的藏品XML總數 ÷ 計畫藏品XML總數 DC十五個元素都計算一次	C類指標為資料完整性品質指標，為雷達圖第三部份。以C.I.1為基礎計算百分比得C部分的分數。 6個必填欄位佔60%，其他9個欄位佔40%。
A.I.1. 元素使用正確性	元素是否正確使用？亦即元素的實際使用情形是否符合每一個Element的定義？如：Format & Type、Subject & Description、Source & Relation等。	人工判斷	A類指標為資料正確性品質指標，為雷達圖第四部份。以A.I.1為基礎計算百分比得A部分的分數 範例：80%



## C.I.1. 元素使用完整性範例

元素名稱[黃底粗體 為必填欄位]	使用元素的藏品數	計畫有使用Title的藏品XML總數 ÷ 計畫藏品XML總數	資料完整性品質指標
<b>TITLE</b>	1280	100%	6個必填欄位佔60%，其他9個欄位佔40%。
CREATOR	1225	95.70%	
<b>SUBJECT</b>	1280	100%	
DESCRIPTION	1280	100%	
<b>PUBLISHER</b>	25	1.95%	
CONTRIBUTOR	0	0.00%	
DATE	1200	93.75%	範例：61.95%
TYPE	1280	100%	
<b>FORMAT</b>	0	0.00%	
<b>IDENTIFIER</b>	1280	100%	
SOURCE	0	0.00%	
LANGUAGE	0	0.00%	
RELATION	0	0.00%	
COVERAGE	1280	100%	
<b>RIGHTS</b>	1280	100%	



## 元素使用正確性評分

- A. 資料對應正確，資料內容呈現一致且良好；
  - B. 資料對應正確，但資料內容呈現有缺失，如呈現方式雜亂不易閱讀、欄位有對應但無資料值等。若歸納至此分級則評等者必須加入評注，說明詳細的缺失情形及建議的改善方式。
  - C. 資料對不正確，但實際資料內容品質良好或重要（如資料正確卻匯錯欄位，做調整後即可成為良好的後設資料）。若歸納至此分級則評等者必須加入評注，說明詳細的缺失情形及建議的改善方式。
  - D. 資料對應不正確，且資料內容品質不良有缺失。
- 元素使用正確性: A-90|B-80|C-70|D-60



## 後設資料的品質指標<sup>4</sup>

- 經由XML記錄的Hyperlink為依據，連回原系統的記錄，包括：

指標	定義	算式	範例	權重
<b>Ac.I.1.</b> a. Hyperlink存在比例	XML中提供的 <Hyperlink>總數 ÷XML記錄總數	計畫原始資料庫 連結總數 ÷ 計畫 藏品XML總數	1280 ÷1280= 100%	(後述)
b. Hyperlink有效比例	XML中提供的 <Hyperlink>有效 總數÷記錄總數	計畫原始資料庫 有效連結總數 ÷ 計畫藏品XML總 數	1280 ÷1280= 100%	
c. Hyperlink相異比例	XML中提供的 <Hyperlink>不重 複總數（扣除連 結至同一頁面）÷ 記錄總數	（計畫原始資料 庫有效連結總數 －有效但重複連 結數）÷ 計畫藏 品XML總數	(1280- 0)÷128 =100%	



## 後設資料的品質指標<sup>5</sup>

指標	定義	算式	範例	權重
<b>Ac.I.2.</b> 數位檔顯示 a.數位檔存在比例	XML中提供的數位檔連 結（<ICON>與 <Multimedia>）總數 ÷ XML記錄總數	計畫提供數位檔連結總 數 ÷ 計畫藏品XML總數 ex：100筆資料提供150 個數位檔URL ÷ 總共100 筆XML	1280 ÷1280= 100%	(後述)
b.數位檔有效比例	XML中提供的數位檔有 效連結總數 ÷ 記錄總數	提供的數位檔有效連結 總數 ÷ 計畫藏品XML總數 ex：150個連結中共有 120個連結有效 ÷ 總共 100筆XML	1280 ÷1280= 100%	
c.數位檔相異比例	XML中提供的數位檔連 結不重複總數（扣除多 筆資料使用同一個數位 檔連結）÷ 記錄總數	（數位檔有效連結總數 －數位檔有效但重複連 結數）÷ 計畫藏品XML 總數 ex：（120個數位檔有效 連結－30個重複連結） ÷ 總共100筆XML	(1280- 0)÷1280 =100%	



## 後設資料的品質指標<sup>6</sup>

指標	定義	算式	範例	權重
d. 具有數位檔有效連結的記錄數	包含有效數位檔的記錄數，不論一筆XML提供幾個有效數位檔皆記為1	計畫提供的藏品XML中，含有數位檔有效連結的XML總數 ex：計畫提供100資料中，有80筆提供了120個數位檔有效連結，記為80	1280	相異Hyperlink之比例。 數位檔部分的指標計算方式為：先以數位檔存在比例（Ac.I.2 a）乘以數位檔有效比例（Ac.I.2 b），再乘以數位檔相異比例（Ac.I.2 c）得有效又相異數位檔之比例。可再視需要以有效數位檔記錄數（Ac.I.2 d）結果對得分微調之。  Ac類指標得分由Hyperlink及數位檔兩部分之得分平均求得。
e. 具有數位檔有效連結的記錄比例	包含有效數位檔的記錄數佔確實新增匯入數的比例	含有數位檔有效連結的藏品XML總數 ÷ 計畫藏品XML總數 ex：承上，計畫提供80筆包含數位檔有效連結的XML ÷ 總共100筆XML	1280 ÷ 1280 = 100%	



## 資料匯出的必備欄位

- 共包括 Title、Publisher、Subject、Identifier、Format 與 Rights



## 相關配套措施-對照服務

- 對照機制的建立與對照表的調整（MAAT與聯合目錄）
  - MAAT小組亦可提供有關標準的應用及對照等規範指引及其訓練推廣課程
  - 各計畫在匯入後設資料至聯合目錄前，由MAAT小組同仁以品質指標進行事前品質控制
  - 後設資料匯入後由聯合目錄團隊進行資料匯入的檢查及整體效益評估
- MAAT→實務規劃→聯合目錄匯入協助
  - <http://metadata.teldap.tw/design/design-frame.html>



## 資料對照的範例

聯合目錄欄位	範例一	範例二	範例三
標題	篇名:帝國領有前に於ける臺灣植物の調査	女用銀鑲金雙股耳挖髮簪	莢社詩稿(第七冊)(數位財產編號ts_a_008_0_001)
資料識別	索書號:505/6438	登錄號:NP0311	數位財產編號:ts_a_008_0_001



## 資料對照的範例

聯合目錄欄位	範例一	範例二	範例三
資料類型	文字	文獻（館訊與論文）	 <p>藏品類型:手稿 資料類型:原件 藏品層次:合集</p>
著作者	作者:川上瀧彌	莊文星	編者不詳



## 資料對照的範例

聯合目錄欄位	範例一	範例二	範例三
貢獻者	謄稿者不詳	相關人物:學生家長 照片人物:學生家長	
語言	漢語		
涵蓋範圍		拍照地點:美濃地區 相關地點:高雄縣美濃鎮	台灣西部 拍攝地點:澎湖縣七美鄉 拍攝地經度:119.2644 拍攝地緯度:23.1314





## 資料對照的範例

聯合目錄欄位	範例一	範例二	範例三
主題與關鍵字：	主題分類:植物—高等植物	澎湖;玄武岩;低平火口;瑪珥火山;火山頸;火山筒;地景;成因探討;台灣	客家六堆
出版者：	出版單位:臺灣博物學會	經濟部中央地質調查所	[未填寫]
日期：	出版年份:明治四十四年(西元1911)	拍攝時間:2006/5/26	拍照年/月:民國60-70年代 入藏時間:民國97年11月



## 資料對照的範例

聯合目錄欄位	範例一	範例二	範例三
格式：	媒體類型:紙本	數位檔案格式:JPEG	原稿一頁
來源：	文獻名稱:臺灣博物學會會報 卷期:v01 號次:n001 所在頁數:p.001-p.008		





## 資料對照的範例

聯合目錄欄位	範例一	範例二	範例三
描述：	髮簪為頭髮梳好後再於髮束邊緣插上鈿之飾物,又會配上以真髮或其他物料的髮絲編成的假髻(通常為高髻),再插上簪、釵、珠花、頂花、步搖等頭飾。... 可用作製造假髮的材料有很多,不同時代、地區會用不同的材料。材質:銅製 功能:裝飾用 照片檔名:P312	摘要:《茨社詩稿》第七冊封面(文/申惠豐) 藏品狀況:泛黃	七美石滬,七美雙心石滬為雙滬房滬,是七美的地標,全國獨一無二。



## 資料對照的範例

聯合目錄欄位	範例一	範例二	範例三
管理權：	數位化執行單位:中華民國行政院農業委員會林業試驗所 典藏單位:林業試驗所圖書館	著作財產權人:屏東縣客家事務處	數位化執行單位:靜宜大學 台灣文學系



## 後設資料品質指標的實例說明

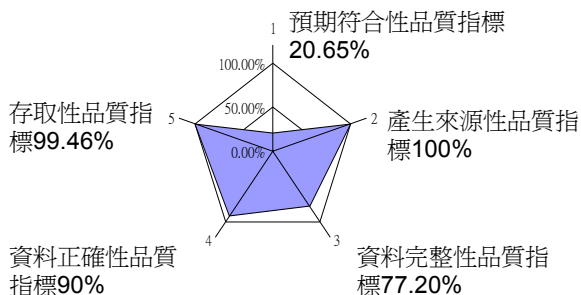
來源-標準規範評分

評分: A-100

計畫於「資料類型」欄位有採取DCMI的控制詞彙規範, 且其他欄位的填寫方式都有其自訂之規範(如「格式」、「主題與關鍵字」)且內容填寫一致, 故給予其100分。

元素使用正確性

評分: A-90



## 後設資料品質指標的實例說明

來源-標準規範評分

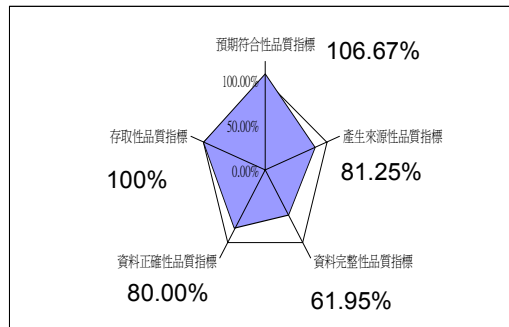
評分: C-75

1. 必填欄位部分, 皆未填寫「格式」欄位, 照片檔案部分, 未填寫「出版者」欄位。
2. 於「日期」、「範圍」與「型式」欄位皆發現有使用自訂的著錄規範。

元素使用正確性

評分: B-80

1. 請補充必填之「格式」與「出版者」欄位資訊。「格式」欄位部分因描述實體為文獻與照片, 其材質皆為紙, 故建議可統一以「材質: 紙」著錄。照片部分如係由XXXXX館出資委託拍攝, 「出版者」欄位建議可統一以「XXXXX館」著錄, 如係個人拍攝出版則建議以作者「OOO」著錄。



2. 建議於「資料類型」欄位加入DCMI的控制詞彙規範, 文獻部分以「型式: 文字、圖像」著錄, 照片部分以「型式: 圖像」著錄。



# 後設資料品質指標的實例說明

## 來源-標準規範評分

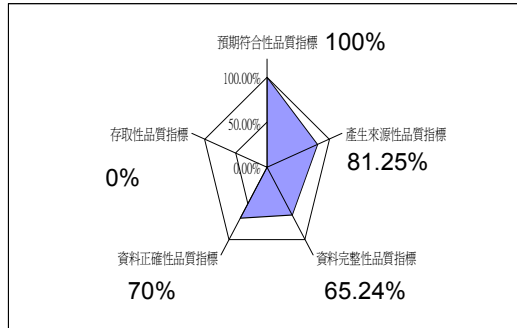
評分: C-75

1. 沒有符合DC標準“**One-to-One Principle**”的原則，若是描述實體照片(材質:紙)，則「格式」應為照片的長寬(不應為數位檔案JPEG)，「日期」應為拍照日期(不應為入藏日期)。
2. 六個必填欄位，缺少「**publisher**」。
3. 部份欄位有使用自訂的著錄規範，如「主題」，「日期都採民國紀年」。

## 元素使用正確性

評分: C-70

1. 資料對應不正確，不符合DC “**One-to-One Principle**”的原則，若是描述實體照片，則「格式」應為照片的長寬(不應為數位檔案JPEG)，「日期」應為拍照日期(不應為入藏日期)。



附註：因結案較久，連結也因當時聯繫問題而付之闕如，故存取性品質指標分數缺乏，



# 後設資料品質指標的實例說明

## 來源-標準規範評分

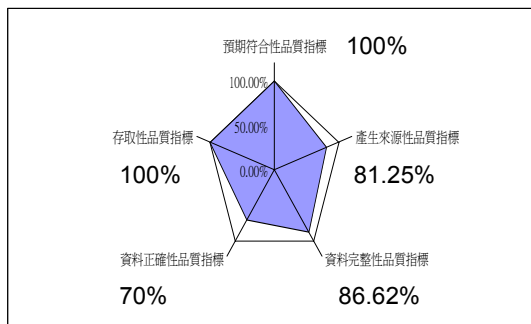
評分: C-75

計畫於「資料類型」欄位有採取DCMI的控制詞彙規範，且其他欄位的填寫方式都有其自訂之規範(如「日期」|「主題與關鍵字」)且內容填寫一致，但典藏品間關係脈絡已消失，故給予75分。應作一筆collection record

## 元素使用正確性

評分: C-70

雖資料類型：「藏品層次-合集」，但實際為單頁的數位圖檔，與合集不符。  
抽檢4及5-描述：有參照至其他圖檔，應轉入關聯欄位





## 問題與討論



## 製作群

數位典藏與數位學習國家型科技計畫 後設  
資料工作組



## 參考書目

- 葉俊宏，98數位典藏專業培訓課程-都柏林核心集與最佳應用規範簡報檔
- 葉俊宏，98數位典藏專業培訓課程-聯合目錄對照原則及程序簡報檔



*The End ....*

歡迎指正

- 後設資料工作組
- 陳慧婷 [winefast@gate.sinica.edu.tw](mailto:winefast@gate.sinica.edu.tw) 02-27899498