

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 陳慧嫻

14 Oct., 2010

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)

Source : <http://catalog.digitalarchives.tw/?URN=1116727>

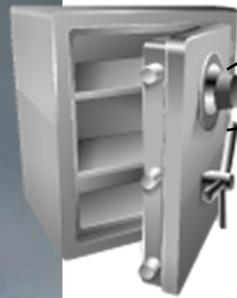
一對一原則

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



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

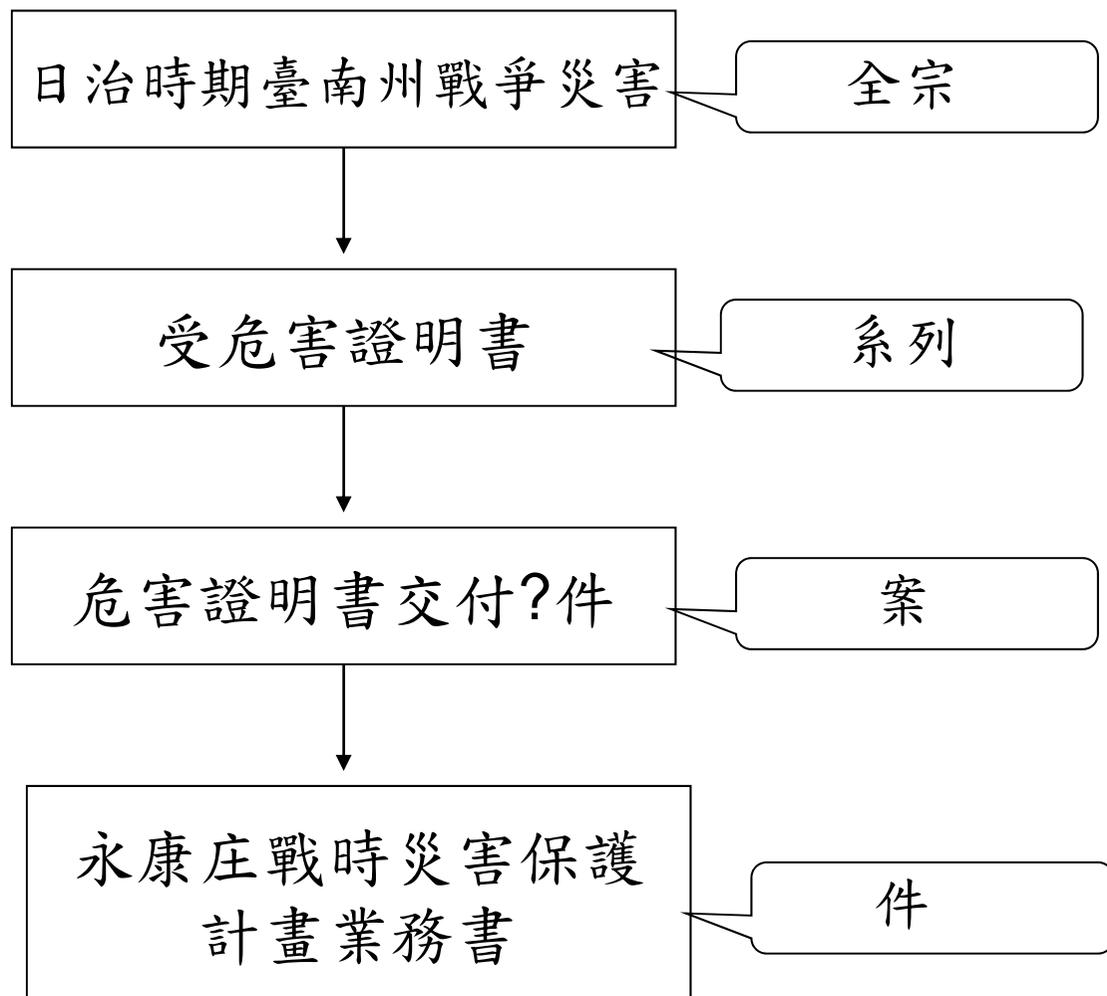
明 青玉四耳活環尊
(故玉005759N000000000) ;
漢 舊玉王 勒
(故玉005760N000000000) ;
... (略, 共47件)

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>

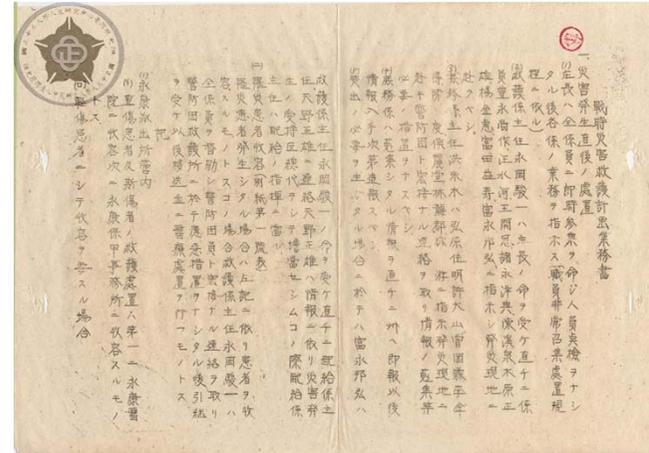
一對一原則



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)

簡化原則

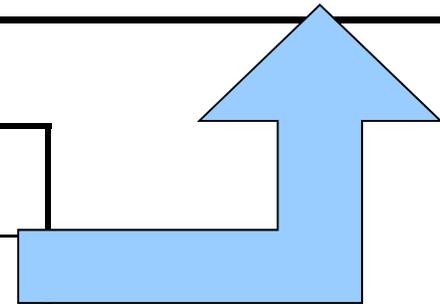
- 目的：保留原有的資訊
- DC欄位中可使用修飾詞（qualifier），或不使用修飾詞
- 建議在匯入資料時，一併匯入計畫原欄位名稱，以資識別資訊。

簡化原則



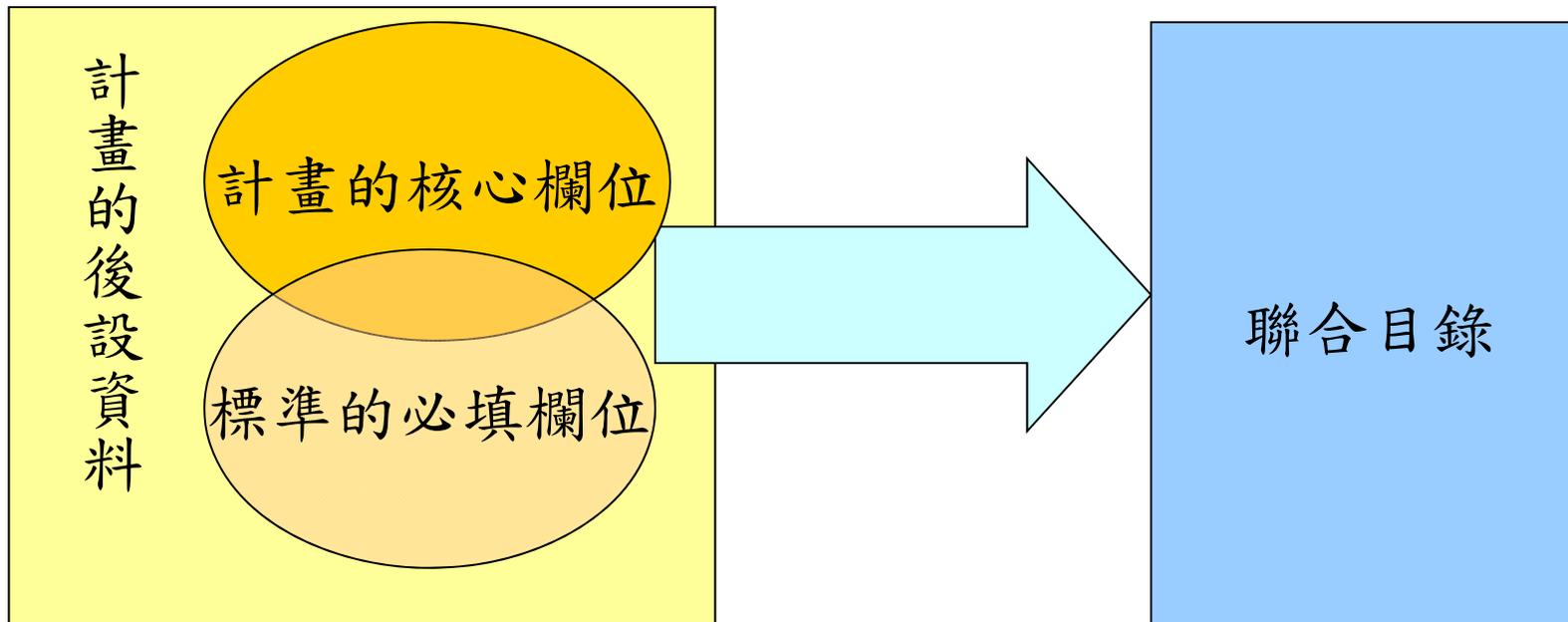
標題	文物名稱：婦女頭飾 英文名稱：female head ornament 族語名稱：vanga
----	--

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



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

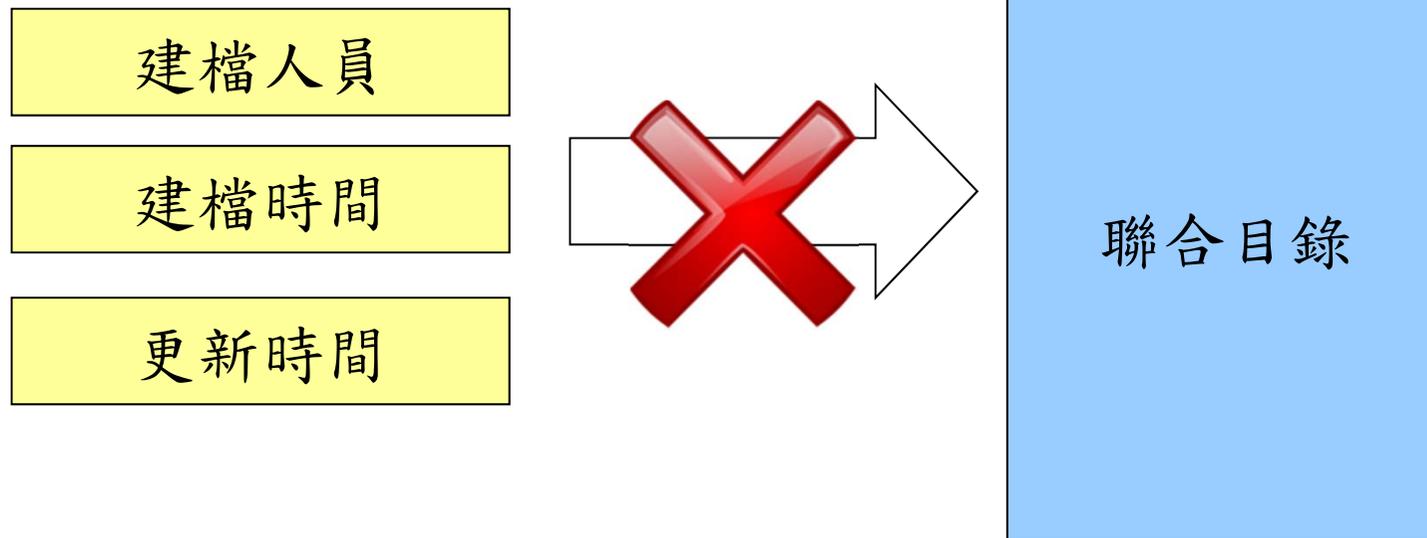
- 若計畫核心欄位或必填欄位不足以滿足聯合目錄必備欄位時
 - 審視其他欄位是否有適合對照至聯合目錄必備欄位者
 - 若原始欄位皆無適合對照至聯合目錄必備欄位時，計畫可增加此欄位資訊，由系統自動或人工匯入

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
----	--

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



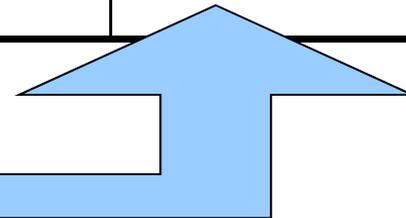
主題	主要主題：拓片 類目：石刻部-祠廟類-佛寺之屬
----	----------------------------

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

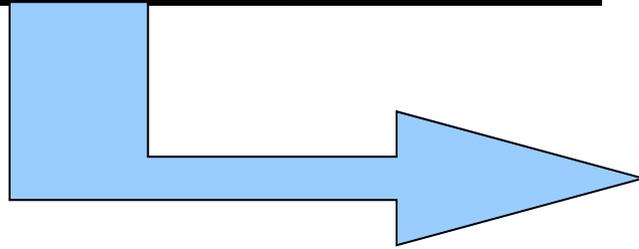
貢獻者	關係人：買主夥記顧岐山、為中人簡鼎宗、知見人男文傑
-----	---------------------------



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

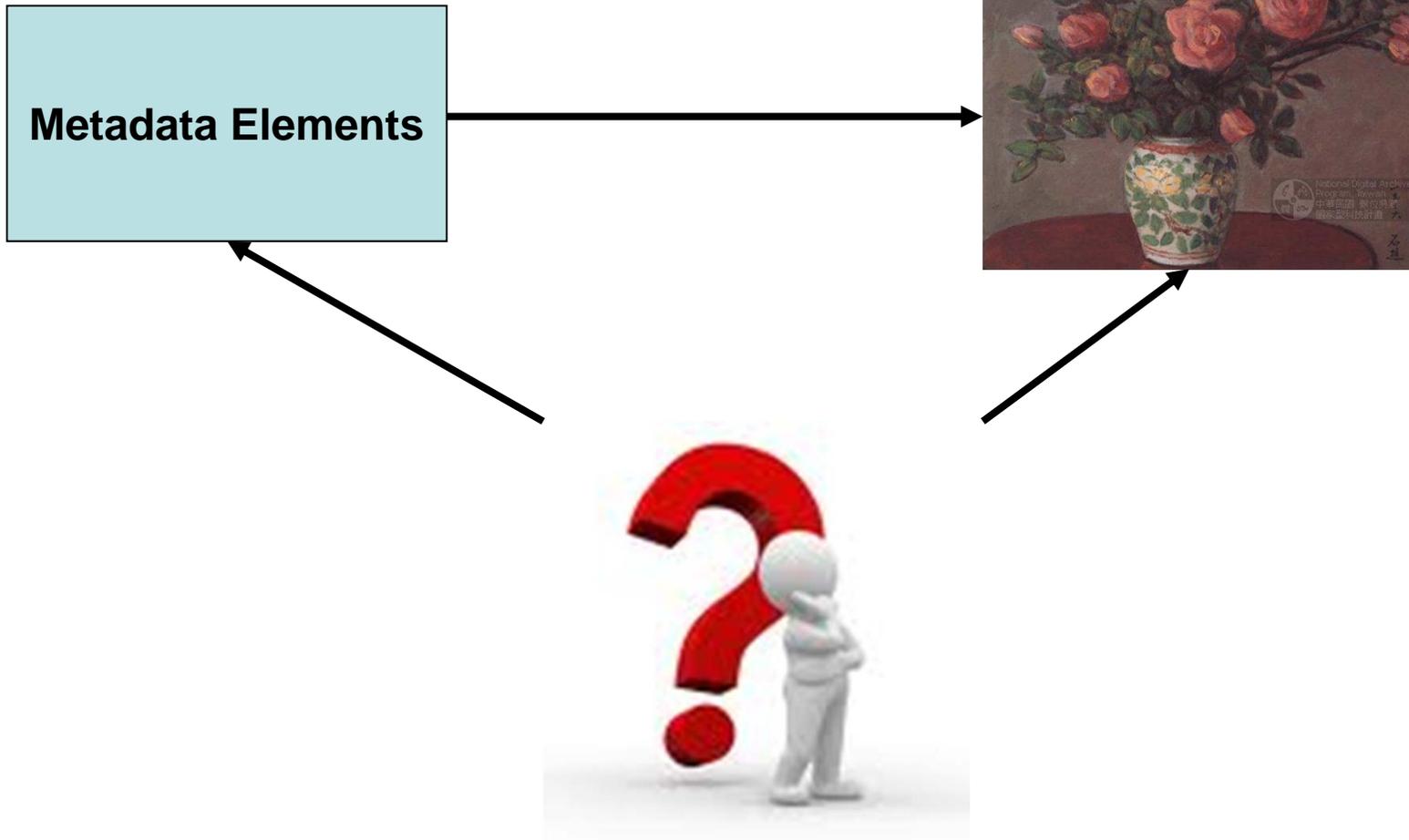
後設資料工作組(MAAT)

陳亞寧 陳慧嫻

大綱

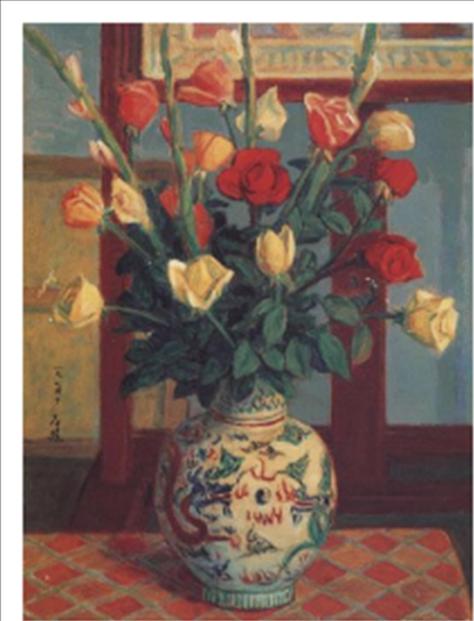
- 說故事的內容重點
- 聯合目錄的資料範例
- 後設資料品質指標的內容簡介
- 資料匯出的必備欄位
- 相關配套措施—對照服務

說故事的內容重點



聯合目錄的資料範例：1974玫瑰花¹

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，而以自然形象所涵蘊之結構為主，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，此外緣圓圈又為葉子所形成之內緣所呼應，花瓶之渾圓就不言而喻了。背景之方塊嵌板顯係蓄意為之，因此方塊與牆面左右不相連屬。桌面自然蘊示方塊，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

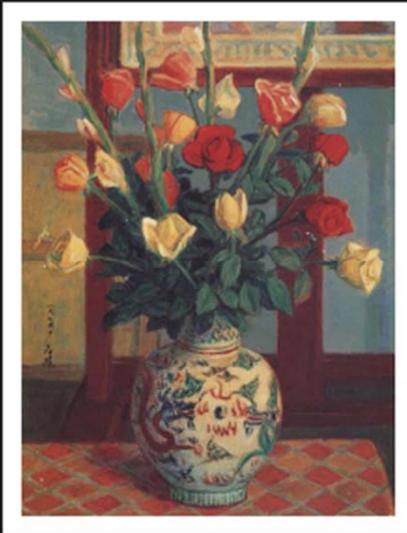
出版者：

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²

1974 玫瑰花



日期：

創作日期：1974

格式：

典藏實體：1件；72.5 x 60.5 cm

關聯：

李石樵著，黃明政主編，林宗興、千普、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

範圍：

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¹

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」 承諾 （於管考報告上填寫的「 實際匯入 」數）新增多少筆藏品XML	計畫當年度承諾匯入之藏品XML總數	734	O類指標為 預期符合性品質指標 ，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：107.77%
O.I.2. 藏品匯入比例	計畫「當年度」 確實 新增匯入筆數佔承諾匯入總數之比例	年度 確實 新增藏品XML總數 ÷ 計畫當年度 承諾 匯入之藏品XML總數	791	

後設資料的品質指標²

指標	定義	算式	範例	權重正規化
P.I.1. 提供單位資訊	Rights or Publisher 欄位是否填值？	Rights or Publisher 欄位非空白的XML總數 ÷ 藏品XML總數	791÷791=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：100%
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	100%	
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	100%	
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	100%	

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³

指標	定義	算式	權重正規化
C.I.1. 元素使用完整性	<p>Title = 計畫有使用Title的藏品XML總數 ÷ 計畫藏品XML總數</p> <p>DC十五個元素都計算一次</p>	<p>Title = 計畫有使用Title的藏品XML總數 ÷ 計畫藏品XML總數</p> <p>DC十五個元素都計算一次</p>	<p>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。</p> <p>6個必填欄位佔60%，其他9個欄位佔40%。</p>
A.I.1. 元素使用正確性	<p>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。</p>	人工判斷	<p>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</p> <p>範例：80%</p>

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
TITLE	791	100%	6個必填欄位佔 60%，其他9個 欄位佔40%。
CREATOR	791	100%	
SUBJECT	791	100%	
DESCRIPTION	791	100%	
PUBLISHER	791	100%	
CONTRIBUTOR	220	27.81%	
DATE	791	100%	範例：87.90%
TYPE	791	100%	
FORMAT	791	100%	
IDENTIFIER	791	100%	
SOURCE	0	0.00%	
LANGUAGE	791	100%	
RELATION	0	0.00%	
COVERAGE	791	100%	
RIGHTS	791	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⁴

- 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
Ac.I.1. 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	791 ÷791=1 00%	(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	791 ÷ 79 1=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	(791- 0) ÷ 791 =100%	

後設資料的品質指標⁵

指標	定義	算式	範例	權重
Ac.I.2.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	791 ÷ 791 = 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	785 ÷ 791 = 99.24%	
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	(785 - 30) ÷ 791 = 99.24%	

後設資料的品質指標⁶

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	791	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。
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	$785 \div 791 = 99.24\%$	Ac類指標得分由Hyperlink及數位檔兩部分之得分平均求得。

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>
 - 授權書

授權書

同意書（範例）

茲同意由「數位典藏與數位學習國家型科技計畫 數位技術研發與整合計畫 系統整合與規範研發計畫 後設資料工作組」（以下簡稱「後設資料工作組」）進行聯合目錄資料匯入後設資料比對服務。並同意依下列各項條款之進行：

- （一）**○○典藏數位化計畫（以下簡稱「○○計畫」）**同意回饋後設資料工作組至少五筆完整範例（含後設資料與其對應之數位檔案），供後設資料工作組援引為範例，作為推廣教育與學術研究之應用。
- （二）授權範圍：所提供之後設資料與數位影像內容之權利，影像檔的著作權（例如以數位相機進行數位化後所得到數位檔案）為**○○計畫**所有，且影像內容（例如原本數位化的對象是照片或畫作）之權利，係經過盤點而權利狀態清楚（包括著作權已經期滿、計畫本身擁有著作權，或是其原本就不受著作權保護），後設資料工作組得於後設資料之教育訓練、推廣、論文發表與出版（含傳統紙本式及網路下載與瀏覽）中運用。
- （三）所引用之各圖片將註明圖片來源計畫，並附註資料上傳於《數位典藏聯合目錄》之網址。
- （四）上述第二款之授權範圍與利用方式，包括為達到後設資料之教育訓練、推廣、教學、論文發表與出版之目的所需要之各種利用行為，例如：重製、公開展示、公開發表、公開傳輸、文件免費下載等方式，出版形式包含紙本、電子形式之出版品及網頁等。
- （五）各項文件均以免費方式流通。若遇以販售方式出版發行的狀況將另行請求授權。
- （六）授權內容：本計畫之承辦人員會另行與**○○計畫**進行協調，選擇五筆範例。
- （七）交付日期及交付方式：由**○○計畫**主動提供，或由後設資料工作組自聯合目錄下載圖檔直接使用。
- （八）上述條款如有重大變更，須另行以書面由雙方共同同意之。

此致「數位典藏與數位學習國家型科技計畫 數位技術研發與整合計畫 系統整合與規範研發計畫」計畫主持人 中央研究院資訊科學研究所與資訊科技創新研究中心合聘研究員 何建明先生

立書同意人

計畫名稱：**中研院○○數位典藏計畫**

計畫主持人：○○○_____（簽名）

所屬機構/單位：**中央研究院資訊科技創新中心**

身分證字號：_____

連絡電話：**(02) XXXX-XXXX**

連絡地址：**台北市研究院路 2 段 128 號 中研院資創中心**

中 華 民 國 _____ 年 _____ 月 _____ 日

後設資料品質指標的實例說明

後設資料工作組(MAAT)

陳亞寧 陳慧婷

聯合目錄

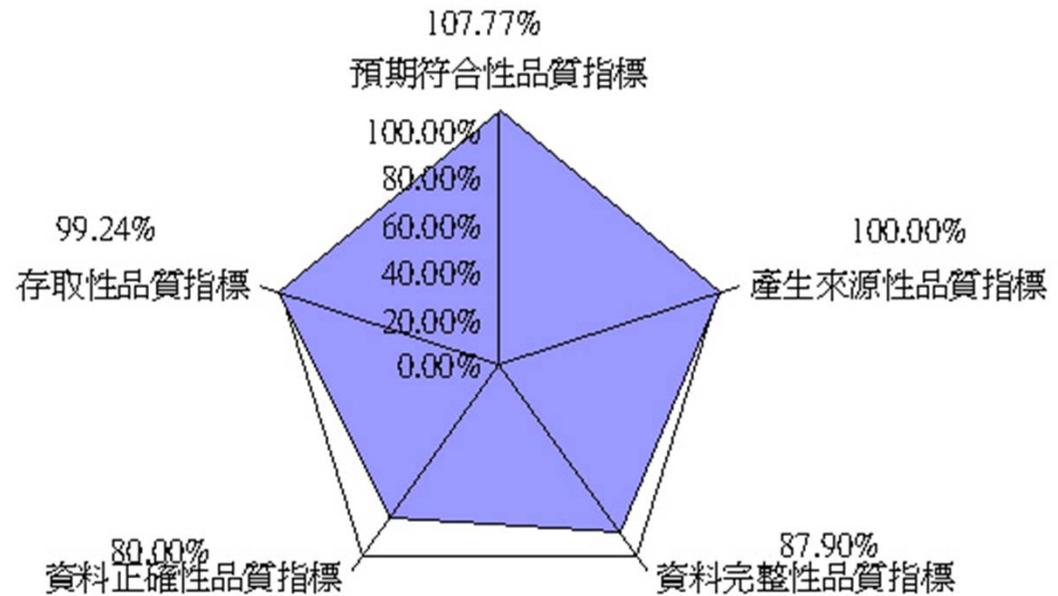
溫淳雅

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**

元素使用正確性
評分: **B-80**

1. 資料對應正確，但資料內容呈現有缺失，
(1)主題與關鍵字欄位同時轉入多個欄位的資料，卻未帶欄位名稱，不易了解意義。
(2)描述欄位也轉入多個凡欄位，建議摘要大綱也帶入欄位名稱。
(3)日期欄位，應採用~作為時間範圍的標點符號。



後設資料品質指標的實例說明

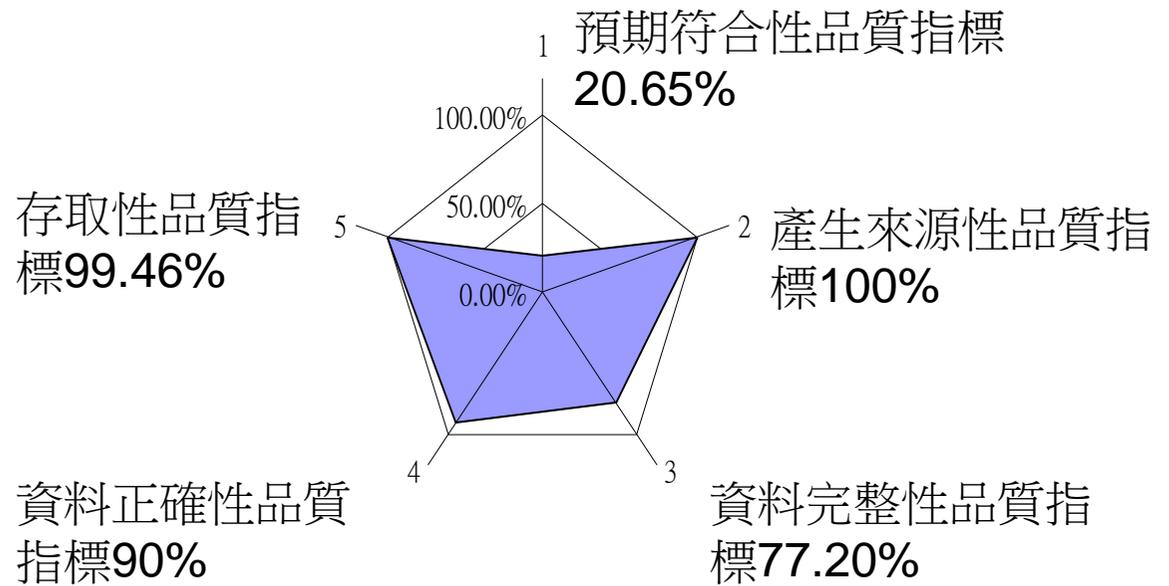
來源-標準規範評分

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: A-100

元素使用正確性

評分: B-80

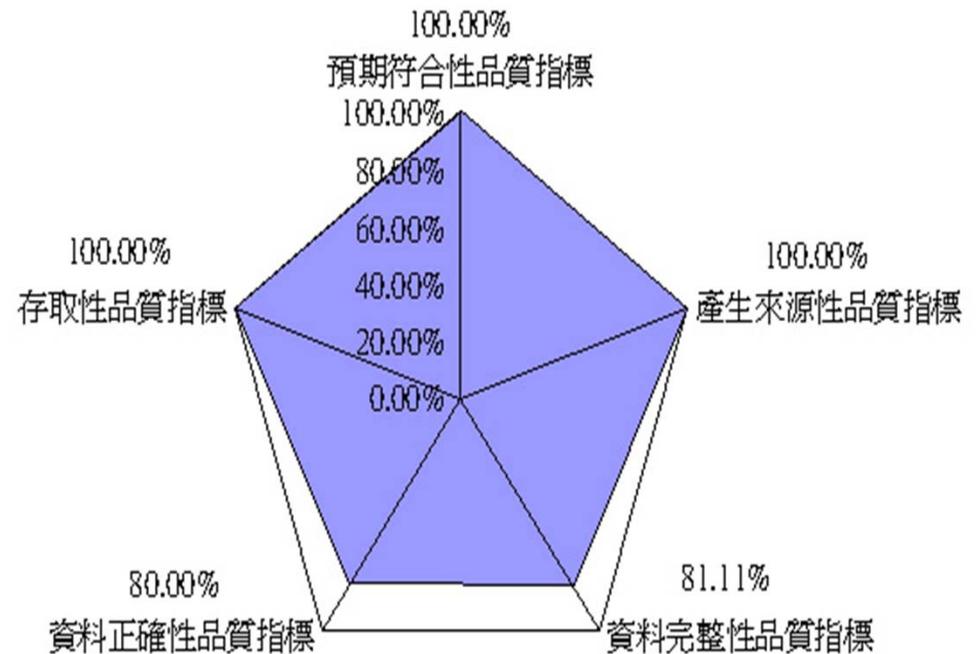
1. 資料對應正確，唯有兩個以上的名稱，應使用相對應之標點符號，如下所示：

(1)貢獻者：有兩個以上之關係人，請以'、'區隔。

(2)範圍：有兩個以上之地名，請以'；'區隔。

(3)著作者：若有兩個以上之著作者，請以'、'區隔。

2. 「貢獻者」、「範圍」、「著作者」：轉出方式不同。根據對照報告建議之規範，有相同欄位名稱，只需出現一次即可。例如：貢獻者→「關係人：」不需重覆著錄。



後設資料品質指標的實例說明

來源-標準規範評分

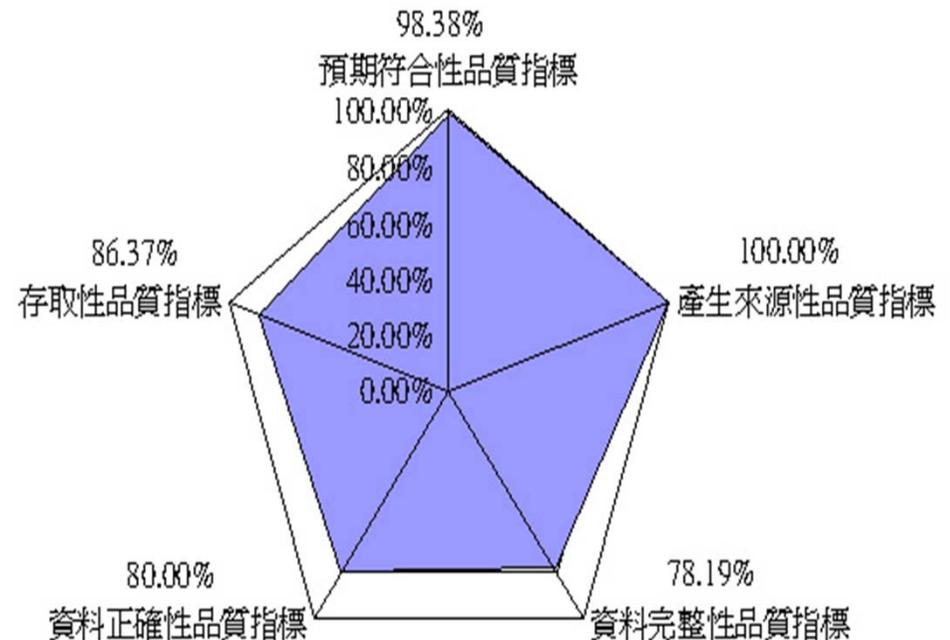
評分: **A-100**

計畫於「資料類型」欄位有採用DCMI Type詞彙，當描述對象為立體作品，如石雕、金屬、木雕時，請轉入「型式：實體物件」；當描述對象為平面作品，如刺繡、石板畫、畫作時，請轉入「型式：靜態影像」。且其他欄位亦皆有按照建議格式編錄，故給予其100分。

元素使用正確性

評分: **B-80**

內容上可以在加強，目前title與subject所填寫的資料值都一樣，若能夠有所區別，對於資料的識別上應更有幫助。



後設資料品質指標的實例說明

來源-標準規範評分

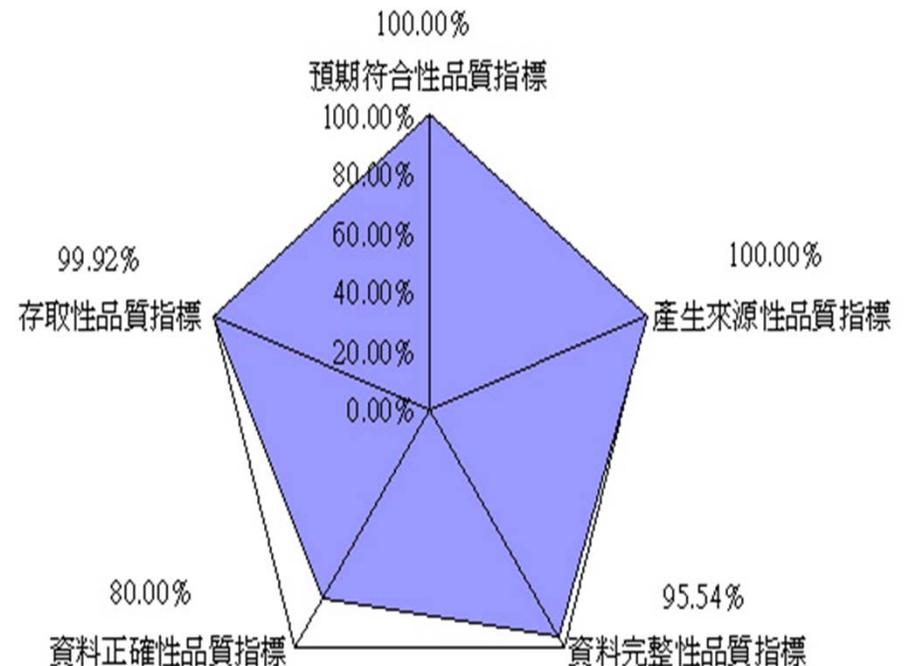
評分: **A-100**

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「標題」|「格式」)且內容填寫一致，故給予其100分。

元素使用正確性

評分: **B-80**

資料對應正確，但資料內容呈現上不易閱讀，如「描述」|「貢獻者」|「關聯」之欄位，應斷行處理(報告書範例中以斷行呈現，但無明確說明，可能導致誤解，所以計畫轉出時都無斷行，導致資料閱讀上有困難)。



後設資料品質指標的實例說明

來源-標準規範評分

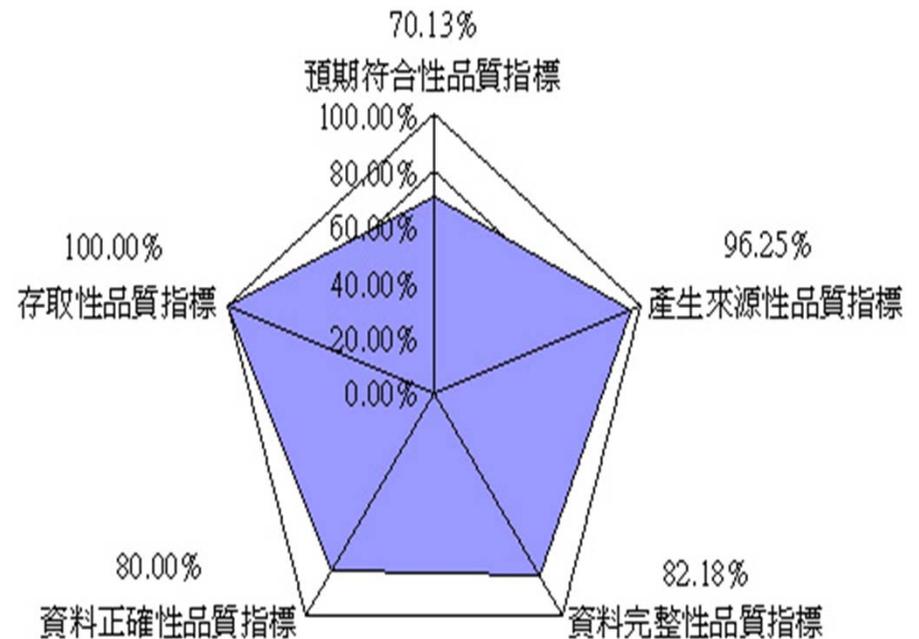
評分: **B-95**

計畫自行定義欄位，資料值符合欄位定義，且大多數的欄位都詳填資料值，且水準一致。

元素使用正確性

評分: **B-80**

關於「撥放時間長度」一詞，計畫網站顯示的欄位為「時間長度」。此外，於後設資料工作組的比對報告中，原文件使用的詞彙均為「播放時間長度」。



問題與討論

製作群

數位典藏與數位學習國家型科技計畫

數位核心平台聯合目錄系統建置子計畫

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 02-27899498