

後設資料品質指標的發展與應用

陳亞寧 鍾豐謙

後設資料工作組 (MAAT)

3 Sep., 2009

Metadata相關辭彙與定義¹

- 相關詞彙
 - 元資料、元數據、詮釋資料、後設資料
 - Catalog, registration, finding aids
 - Meta-data, metadata record
- 定義
 - Data about data.
 - Data about other data.
 - Structured data about data.
 - It most commonly refers to descriptive information about web resources.

Metadata相關辭彙與定義²

- Metadata record
 - 由一組必要的屬性或元素所組成，以描述資源（resource）。
 - 所謂的資源
 - 係指電子文獻或實際的實體物件（an electronic document or a "real" physical object）（DCMI FAQ, 2003）
 - Document-like object or non-document resources（Using Dublin Core, 2005）
 - 具有URL的物件（anything addressable via a URL）（W3C）
 - 與原始物件的關係
 - Metadata 紀錄獨立存在
 - 內嵌在原始物件之內

Metadata實例及獨立存在

清 錦地團花蒔繪三層小屨盒



後設資料

授權與引用資訊

連結到原始資料網站

資料識別：

故玉004103D0000001000

資料類型：

木竹漆器

描述：

二屨及最上層格內置各式小玉器，盒外黑漆地上均以平蒔繪菱格形梅花紋為地紋，其上高蒔繪折枝花卉團花，團花內地文或平目地或梨地，溶合多種技法，十分華麗。中層屨平蒔繪的技法和其他地方明顯不同，可能是清宮修補。

日期：

1644 A.D.-1911 A.D.

格式：

長8.3公分 寬7.8公分 高17.7公分

關聯：

《清宮蒔繪—院藏日本漆器特展》

管理權：

國立故宮博物院

內嵌於原始物件內



<HTML>

<HEAD>

<TITLE>國立政治大學圖書館</TITLE>

<META HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=big5">

<META NAME="keywords" CONTENT="大學圖書館, 政治大學圖書館, 政大圖書館, 圖書館, 社會科學, 國立政治大學圖書館首頁">

<META NAME="description" CONTENT="政大圖書館為提供良好資訊服務不遺餘力.並以提倡校內閱讀及研究風氣為使命.本館提供多樣化之服務:圖書新書查詢,資料庫檢索,每月專題指引,館際互借合作,線上申請服務">

Metadata功能

- 管理 (administrative)
- 描述 (descriptive)
- 長久保存 (preservation)
- 技術 (technical)
- 使用 (use)



現有 Metadata 標準的種類與現況

- Field 
- Value 
- Format 
- Exchange 

常用的Metadata標準¹

- 著錄的結構：Data fields and structure
- 著錄的規範：Data content
- 著錄的內容值：Data value

常用的Metadata標準²

著錄結構

- Generic MD : DC, Darwin Core, CDWA Lite
- Markup language : EAD, TEI Lite, METS
- Domain
 - MARC21, MODS
 - CDWA, VRA
 - EAD
 - HISPID
 - OLAC

常用的Metadata標準³

著錄規範

- AACR/RDA
- CCO
- ISAD
- DACS

常用的Metadata標準⁴

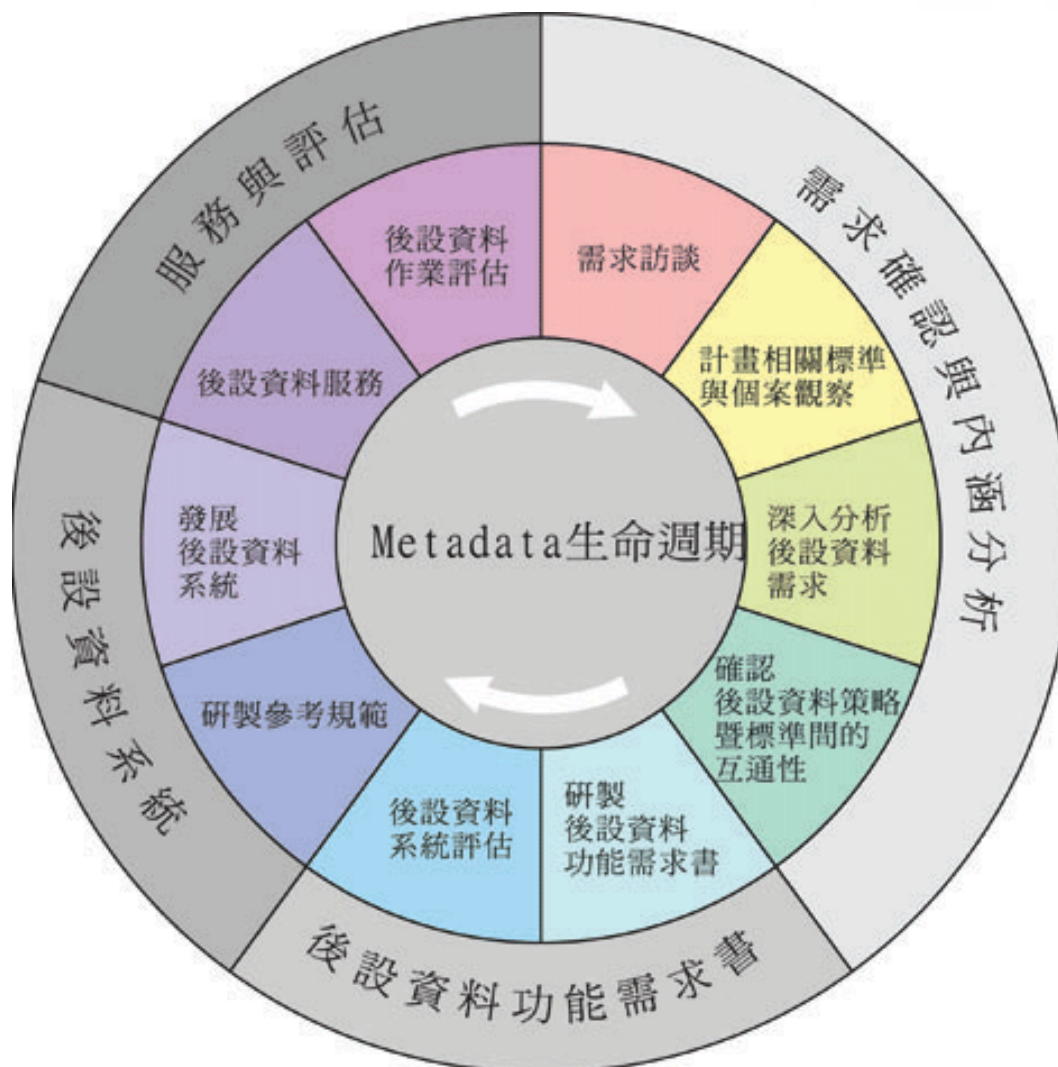
詞彙

- LCSH, LCNAF
- AAT
- TGN
- ULAN
- ICONCLASS
- Species 2000

號碼標記

- DDC
- LCC

後設資料生命週期



MAAT服務項目

- 需求分析
- 內涵結構分析
- 標準分析與建議
- 國際標準接軌與對照
- 需求規格書
- 系統/工具測試與建議
- 著錄規範
- 應用評估

聯合目錄資料範例：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

Metadata品質的提升

- 從90年起，在國科會的經費贊助下，有關數位化計畫的發展與成果已有一定的基礎
- 從數位典藏國家型科技計畫（NDAP）至數位典藏與數位學習國家型計畫（TELDAP）發展過程中，也必須從數量目標的追求轉為品質的提升，以利於未來數位內容產業的加值與再利用

目的

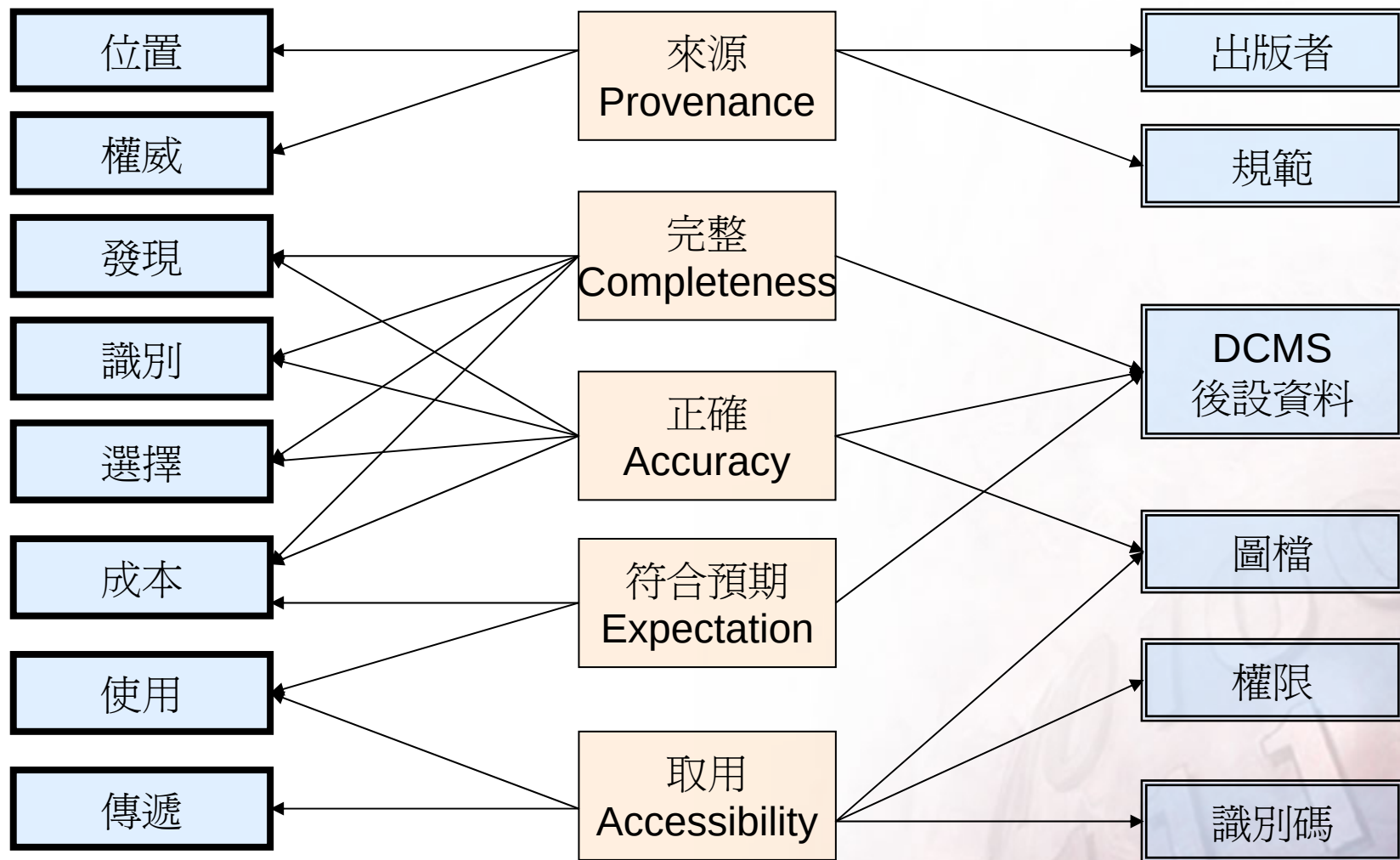
- 主要對象為後設資料記錄，執行記錄內容的品質確保與改善，主要的預計目的有：
 - 提供主題計畫進行自我評鑑
 - 符合國際後設資料標準的要求及其接軌，以利於資源發掘與交換共享
 - TELDAP年度成果評鑑的參考指標
 - 促進資料的再利用與商業加值

後設資料品質指標的依據

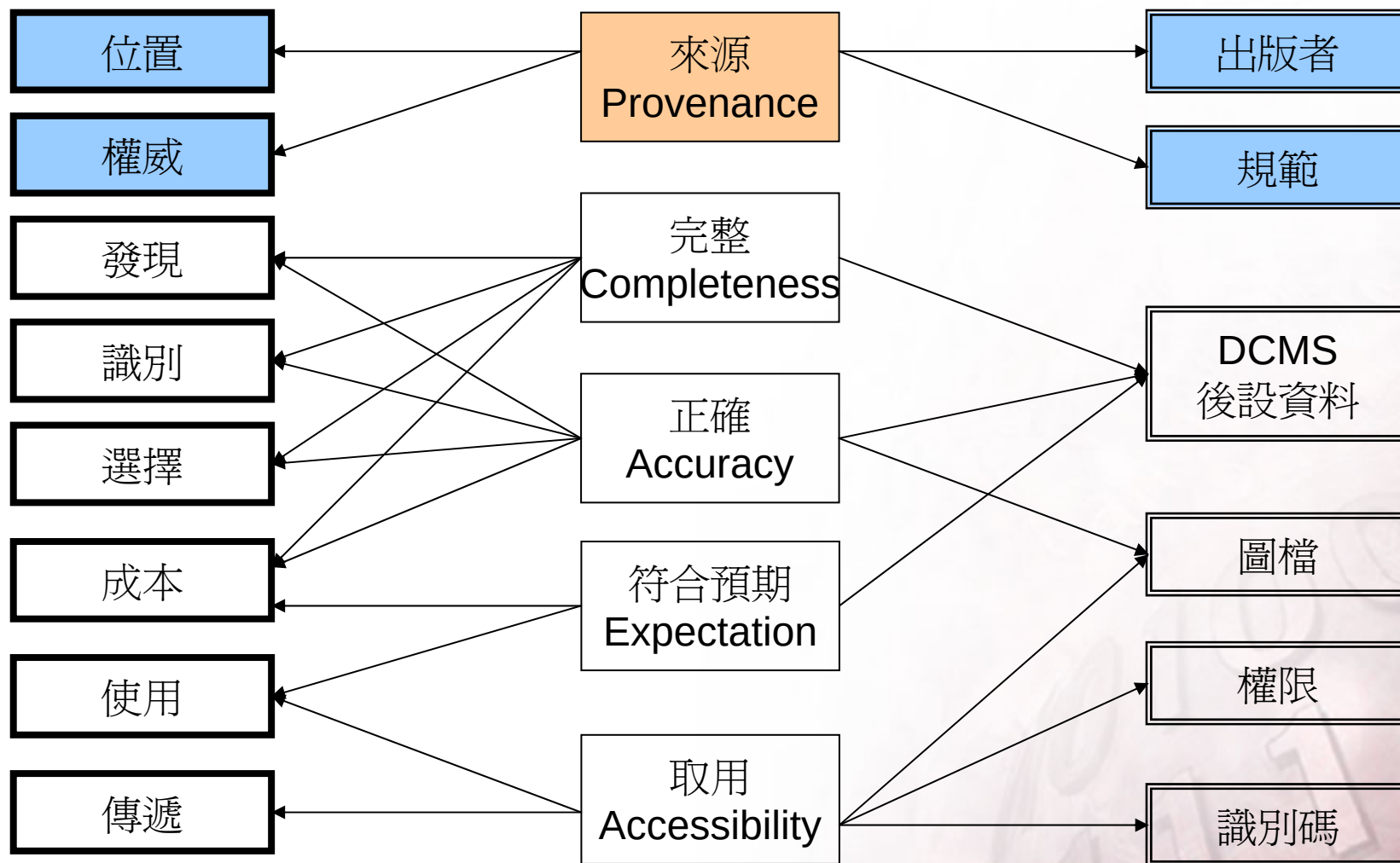
- 預期符合性 (conformance to expectations)
 - 來源性 (provenance)
 - 完整性 (completeness)
 - 正確性 (accuracy)
 - 取用性 (accessibility)
 - 一致性 (logical consistency and coherence)
 - 即時性 (timeliness) : 包括新穎性 (currency) 與時間延遲性 (lag)
- Bruce, T.R. and Hillmann, D.I. (2004) . The continuum of metadata quality: Defining, expressing and exploiting. In Metadata in Practice (pp. 238-256) , edited by D.I. Hillmann and E.L. Westbrooks (Chicago: American Library Association) .



後設資料品質指標及其後設資料元素與功能關係圖

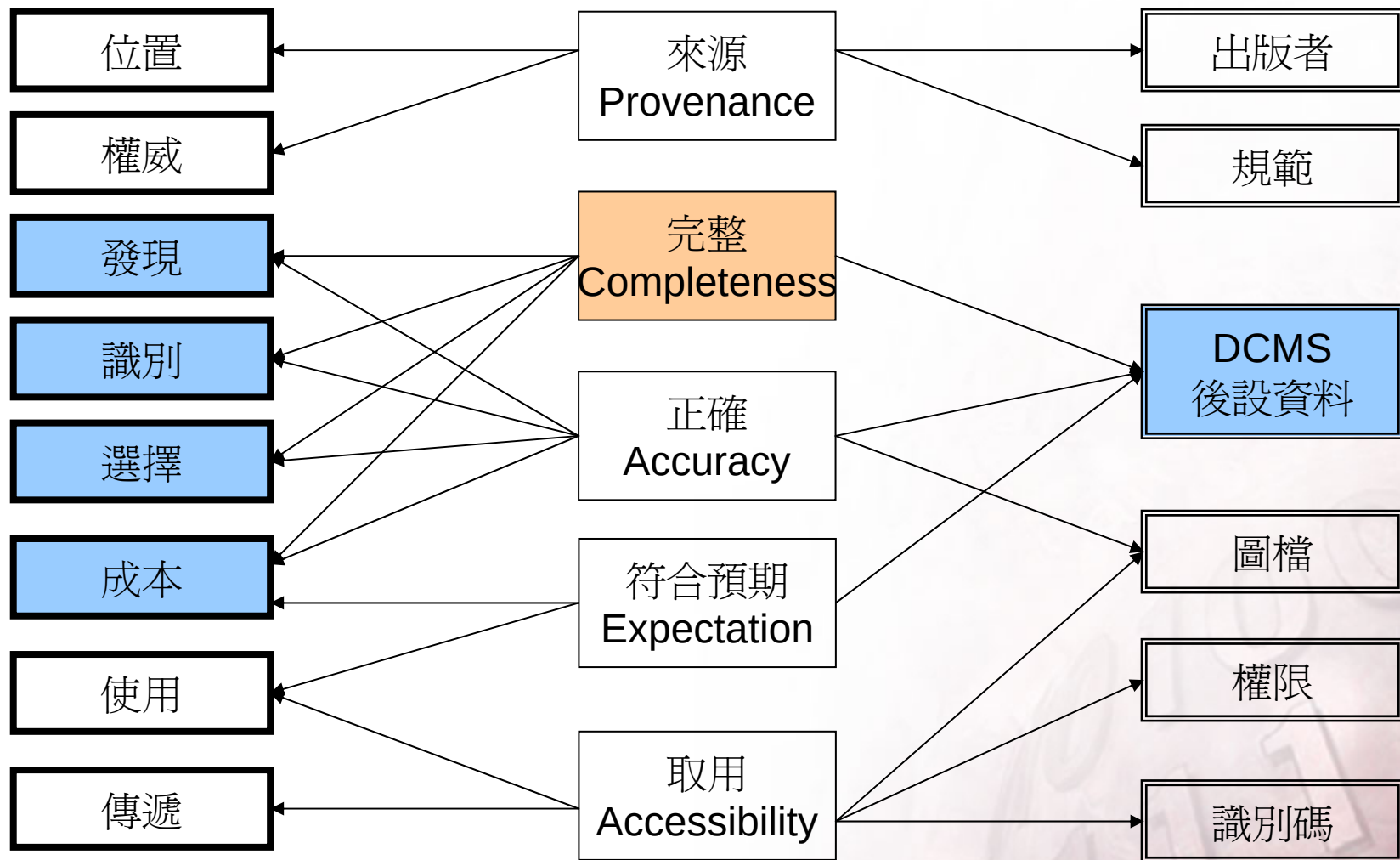


後設資料品質指標及其後設資料元素與功能關係圖



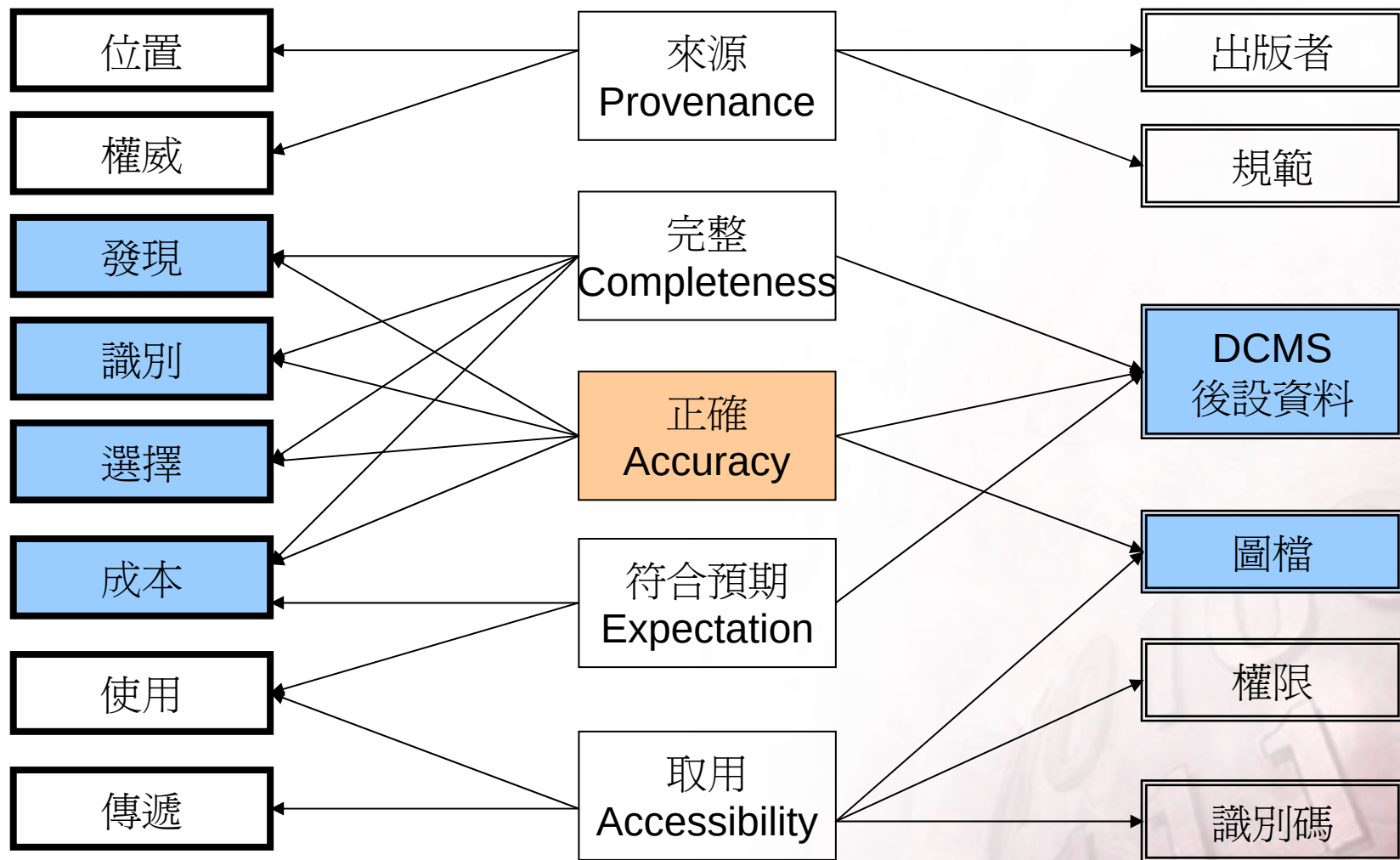


後設資料品質指標及其後設資料元素與功能關係圖

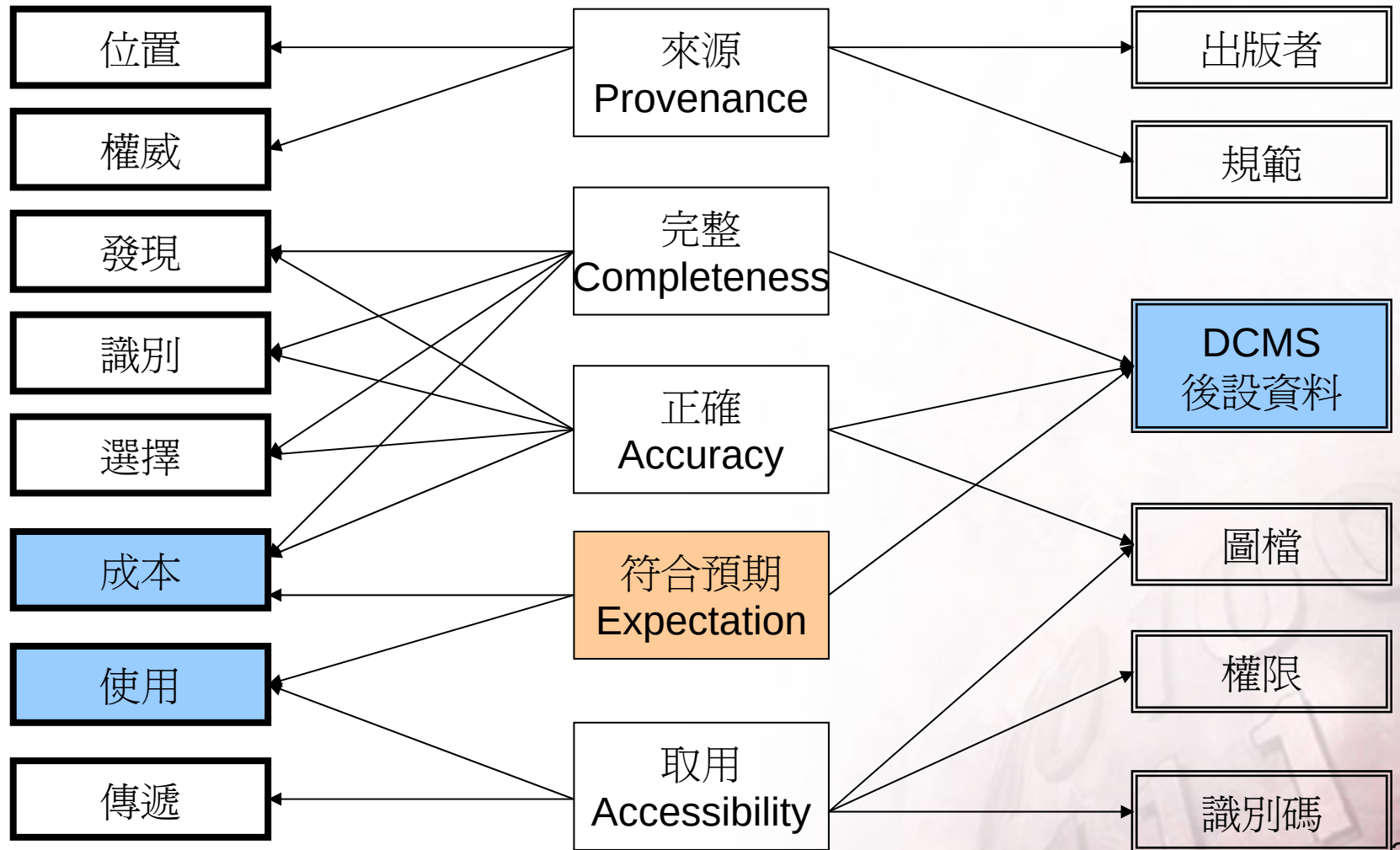




後設資料品質指標及其後設資料元素與功能關係圖

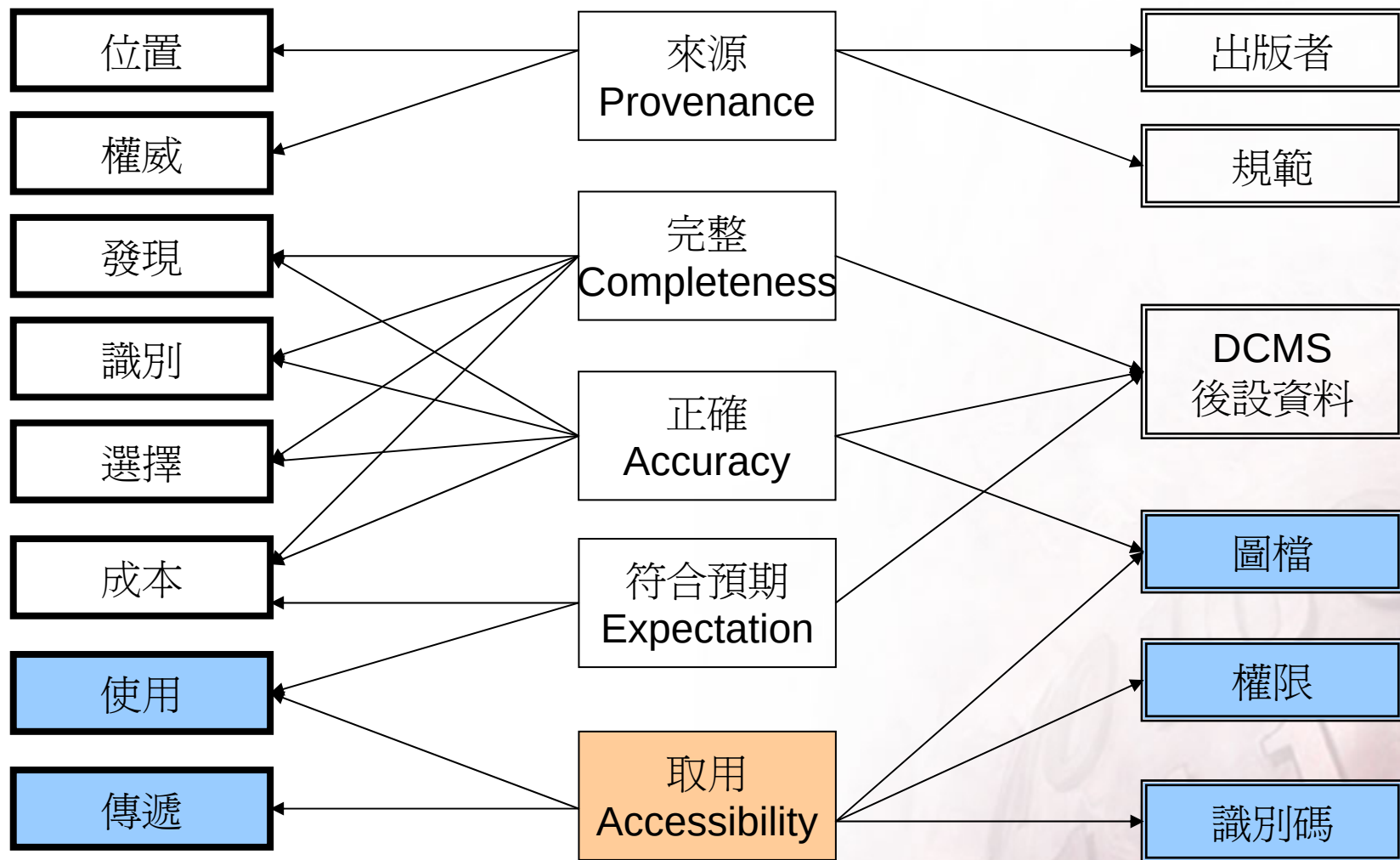


後設資料品質指標及其後設資料元素與功能關係圖





後設資料品質指標及其後設資料元素與功能關係圖

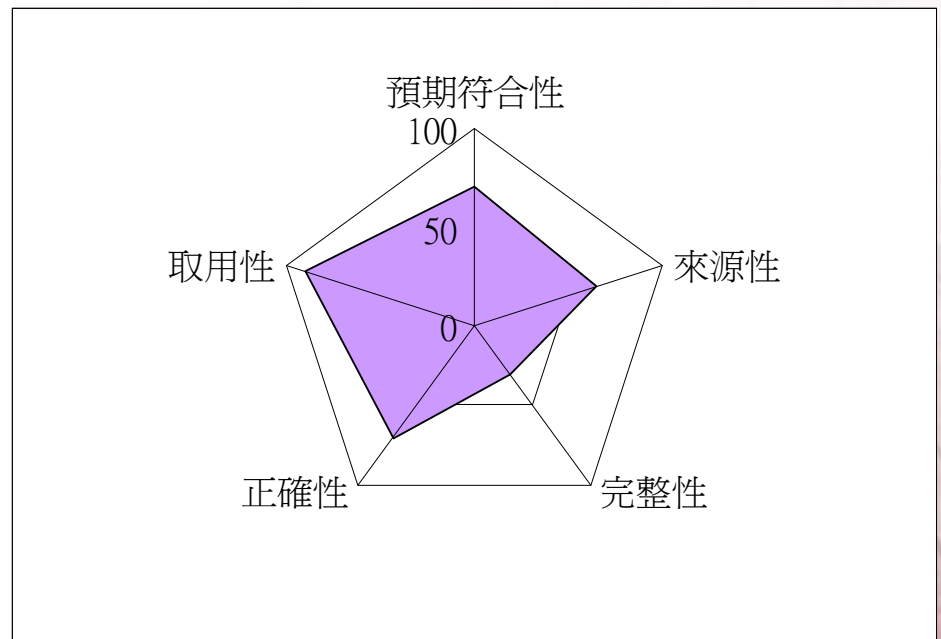


後設資料品質指標

種類	說明	檢核要點
預期符合性	後設資料及藏品數位檔的產出數量是否符合預期目的	藏品數量
		藏品匯入比例
		數位檔／紀錄匯入比例
來源性	後設資料的來源與產生方式是否一致且有控制	提供單位的資訊
		記錄元素的著錄規範
		控制詞彙、索引典或權威檔的採用
		記錄元素資料值的著錄規範
完整性	後設資料的著錄是否完整	DC 15個元素使用的完整性
正確性	後設資料的著錄是否正確	DC 15個元素使用的正確性
存取性	後設資料及藏品數位檔的相關頁面是否皆能正確存取	原始資料庫連結的存在比例
		原始資料庫連結的有效比例
		原始資料庫連結的相異比例
		數位檔的存在總數
		數位檔的有效比例
		數位檔的相異比例

後設資料品質指標計算與結果

- 以權重、平均等方式計算各之指標之得分，並以雷達圖方式呈現
- 分數並非絕對，也非評鑑之唯一方式，僅呈現客觀的觀察結果，並提供未來改善之建議



後設資料品質指標

種類	說明	檢核要點
預期符合性	後設資料及藏品數位檔的產出數量是否符合預期目的	藏品數量
		藏品匯入比例
		數位檔／紀錄匯入比例
來源性	後設資料的來源與產生方式是否一致且有控制	提供單位的資訊
		記錄元素的著錄規範
		控制詞彙、索引典或權威檔的採用
		記錄元素資料值的著錄規範
完整性	後設資料的著錄是否完整	DC 15個元素使用的完整性
正確性	後設資料的著錄是否正確	DC 15個元素使用的正確性
存取性	後設資料及藏品數位檔的相關頁面是否皆能正確存取	原始資料庫連結的存在比例
		原始資料庫連結的有效比例
		原始資料庫連結的相異比例
		數位檔的存在總數
		數位檔的有效比例
		數位檔的相異比例

後設資料的品質指標-預期目的¹

指標	定義	算式	權重正規化
藏品數量	計畫「當年度」實際新增多少筆藏品XML	計畫當年度實際新增匯入之藏品XML record總數	預期符合性品質指標為雷達圖第一部份。以年度新增匯入record總數與計畫實際達成的record數相除求取提交率得預期符合性品質指標的分數，再視數位檔提供的多寡予以調整
藏品匯入比例	計畫「當年度」確實新增匯入筆數佔實際產出匯入總數之比例	年度確實新增匯入record總數 / 計畫當年度實際產出之藏品XML record總數	
數位檔／紀錄匯入比例	平均每筆紀錄包含的數位檔數量，若提供數位檔越多則此項得分越高	數位檔匯入總數 / 計畫當年度確實新增匯入之藏品XML record總數	

後設資料的品質指標-預期目的²

中央研究院

民族學研究所 生物多樣性研究中心 近代史
研究所 計算中心 臺灣史研究所 語言學研
究所 歷史語言研究所
(927,785筆)

國史館

國史館典藏國家檔案與總統文物數位化
計畫 國家歷史資料庫 總統文物多元精
探索與深耕計畫
(336,856筆)

國家圖書館

古籍文獻典藏數位化計畫 國家圖書館國家典

國立臺灣大學

人類學系 生態學與演化生物學研究所



後設資料的品質指標-預期目的³

- 例：中研院後設資料工作組98年度產出5000筆後設資料，後實際匯入4800筆，故其匯入比例之得分為 $4800 \div 5000 = 96\%$

後設資料的品質指標-來源¹

指標	定義	算式	權重正規化
提供單位資訊	Publisher或Rights欄位是否填值？	Publisher或Rights欄位非空白的紀錄總數 / 紀錄總數	產生來源性為雷達圖第二部份。四項指標各佔25%，加總得來源性的分數
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	
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？	人工判斷	
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	

後設資料的品質指標-來源³

西周晚期 毛公鼎



後設資料

授權與引用資訊

連結到原始資料網站

資料識別：

中銅000651N0000000000

資料類型：

銅器、金屬器

描述：

口飾重環紋一道，敞口，雙立耳，三蹄足。

日期：

899 B.C.-771 B.C.

格式：

通高53.8公分 腹深27.2公分 口徑47公分

關聯：

《故宮西周金文錄》

《國立故宮博物院藏品選目》

管理權：

國立故宮博物院

後設資料的品質指標-來源²

- 例：中研院後設資料工作組於98年度實際匯入的4800筆後設資料紀錄中，有4320筆記錄於「Publisher」欄位中著錄「中央研究院後設資料工作組」，並於「Rights」欄位中著錄「姓名標示-非商業性-相同方式分享 2.5 台灣」，有480筆Publisher或Rights欄位資料有缺漏
- 故提供單位資訊之得分為 $4320 \div 4800 = 90\%$ ，產生來源性得分考量另外三項人工判斷的指標得分可能為 $(90\% + 100\% + 100\% + 100\%) \div 4 = 97.5\%$

後設資料的品質指標-完整性¹

指標	定義	算式	權重正規化
元素使用完整性	Title = 計畫有使用Title的藏品record總數 / 計畫藏品record總數	DC十五個元素都計算一次	資料完整性為雷達圖第三部份。以元素使用完整性為基礎計算百分比得資料完整性的分數

後設資料的品質指標-完整性³

中文名：籐壺



後設資料 授權與引用資訊 連結到原始資料網站

資料識別：
館號：0000001360.1

資料類型：
分類：物料工具和裝備

主題與關鍵字：
族群：泰雅(Atayal)

描述：
製造方式：編
材質：籐
顏色：深咖啡
功能用途：盛置器具
真實性：原件
其他：貯存及收納用

出版者：
國立自然科學博物館 (<http://www.nmns.edu.tw/>)

格式：
尺寸長：0cm
尺寸寬：46cm
尺寸高：72cm

來源：
自然與人文數位博物館

語言：
中文

範圍：
採集地點：中華民國台灣省花蓮縣光復鄉

管理權：
國立自然科學博物館 (<http://www.nmns.edu.tw/>)

後設資料的品質指標-完整性²

- 例：中研院後設資料工作組於98年度實際匯入的4800筆後設資料紀錄中，每筆紀錄都填寫了標題、關鍵字、描述……等15個Dublin Core中的欄位
- 故元素使用完整性得分為 $15 \times 4800 \div 15 \times 4800 = 100\%$

後設資料的品質指標-正確性¹

指標	定義	算式	權重正規化
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	資料正確性為雷達圖第四部份。以資料正確性品質指標為基礎計算百分比得資料正確性的分數

後設資料的品質指標-正確性³

宋至明 銅蟾蜍硯滴



後設資料 授權與引用資訊 連結到原始資料網站

資料識別：
中銅001449N0000000000

資料類型：
銅器、金屬器

描述：
三足蟾蜍，睜目直視，器身嵌銀絲，背部嵌成一個個小圓圈以示蟾蜍之疣；雙眼後方各沿伸一帶狀突起，腹部兩側亦各有一帶狀突起。背部中央有一矮圓柱，附一銅管，腹中空。

日期：
900 A.D.-1699 A.D.

格式：
全高6.4公分 縱長12.8公分

關聯：
《千禧年宋代文物大展》
《國立故宮博物院—文房聚英》

管理權：
國立故宮博物院

後設資料的品質指標-正確性²

- 例：中研院後設資料工作組於98年度實際匯入的4800筆後設資料紀錄中，每筆紀錄都對照至正確的欄位，並填寫了正確的內容（可透過後設資料對照服務的申請與執行來確保）
- 故元素使用正確性得分為 $15 \times 4800 \div 15 \times 4800 = 100\%$

後設資料的品質指標-存取性¹

- 經由後設資料記錄的原始資料庫連結為依據，連回原系統的記錄，包括：

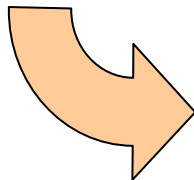
指標	定義	算式	權重
原始資料庫連結存在比例	原始資料庫連結總數÷記錄總數	計畫原始資料庫連結總數 / 計畫藏品總數	(後述)
原始資料庫連結有效比例	原始資料庫連結有效總數÷記錄總數	計畫原始資料庫有效連結總數 / 計畫原始資料庫連結總數	
原始資料庫連結相異比例	原始資料庫連結不重複總數(扣除連結至同一頁面)÷記錄總數	(計畫原始資料庫有效連結總數 - 重複連結總數) / 計畫藏品總數	

後設資料的品質指標-存取性²

北宋 夷則鍾



國立故宮博物院



故宮器物典藏資料檢索

簡目顯示

品名	北宋 夷則鍾
時代	北宋 崇寧四年
尺寸	全高28.2公分 舞脩×舞廣 16.5×13.3公分 鼓間14.9公分 銑 間18.4公分 甬高6.1公分
質材	青銅
功能	樂器
典藏號	中銅000742N0000000000

說明
鐘鈕立雕式龍紋，頸與身的線條轉折趨於裝飾性，尤其龍爪上舉與相向的龍爪成方鉤狀，顯然此套鐘所取法的是南方的裝飾趣味。敲擊(夷則)鐘口的任何一點，所測之音為升C調。與(夷則)相對應的鉦部銘文則由「大晟」而改刻「大和」，書體方正，與原有的圓轉篆書有別。



國立故宮博物院
NATIONAL PALACE MUSEUM

放 大

回首頁

回上一層

重新檢索

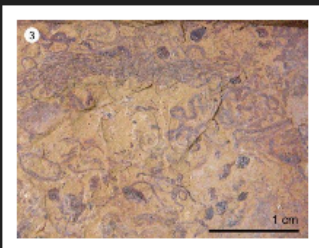
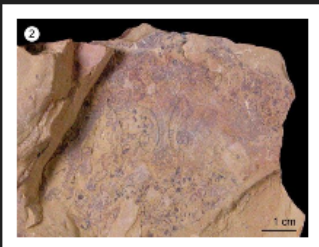
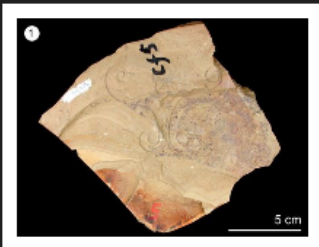


後設資料的品質指標-存取性³

指標	定義	算式	權重
數位檔存在總數	數位檔連結總數	計畫提供的數位檔連結總數 (ex: 有提供70筆圖檔連結)	
數位檔有效比例	數位檔連結有效的比例	數位檔連結有效總數 / 數位檔連結總數 (ex: 70筆圖檔共有65筆圖檔有效)	
數位檔相異比例	扣除多筆連結其實指向同一個數位檔連結者	(有效數位檔連結總數 - 有效數位檔重複的連結總數) / 有效數位檔連結總數 (ex: 60筆圖檔URL不重複 / 65筆圖檔URL有效)	

中文名：中華細絲藻

英文名：*Sinocylindra* sp



後設資料

授權與引用資訊

連結到原始資料網站

資料識別：

館號：NMNS002459 - F029397

資料類型：

澄江生物群化石標本

主題與關鍵字：

英文類群名：Chengjiang biota

中文類群名：澄江生物群

英文類名：Algae

中文類名：藻類

描述：

特徵描述：圖1為部份劈開的泥岩標本，顯示局部有細長藻體密集埋藏的現象，圖2、3分別為不同倍率的放大近觀

地層：筇竹寺組玉案山段（帽天山泥岩）

出版者：

國立自然科學博物館 (<http://www.nmns.edu.tw/>)

日期：

地質年代：寒武紀玉案山組

來源：

自然與人文數位博物館

語言：

中文

範圍：

採集地：雲南省澄江縣帽天山

管理權：

國立自然科學博物館 (<http://www.nmns.edu.tw/>)

後設資料的品質指標-存取性⁵

- 存取性品質指標為雷達圖第五部份，由原始資料庫連結與數位檔存取性兩類指標組成。
- 原始資料庫連結部分的指標計算方式為：先以原始資料庫連結存在比例乘以原始資料庫連結有效比例，再乘以原始資料庫連結相異比例得有效又相異原始資料庫連結之比例。
- 數位檔部分的指標計算方式為：先以數位檔存在總數乘以數位檔有效比例，再乘以數位檔相異比例得有效又相異圖檔之比例。
- 存取性得分由原始資料庫連結及數位檔連結兩部分之得分平均求得。

後設資料的品質指標-存取性⁶

- 例：中研院後設資料工作組於98年度實際匯入的4800筆後設資料紀錄中，共有4800筆原始資料庫聯結，但只有4600筆有效(可連結)，其中又有4320筆為不重複資料，故原始資料庫連結存取性的得分為 $4600/4800 \times 4320/4600 = 90\%$
- 又中研院後設資料工作組於98年度實際匯入的5000筆數位檔中，但只有4825筆有效(可連結)，其中又只有4750筆為不重複資料，故數位檔存取性的得分為 $4825/5000 \times 4750/4825 = 95\%$
- 故存取性品質指標總得分為 $(90\% + 95\%) \div 2 = 92.5\%$

配套措施-指引

- 好的後設資料
 - 符合至少一套國際標準
 - 符合元素的定義
 - 著錄格式一致
 - 必備或基本欄位的著錄要求
 - 反映物件的意義
 - 註明提供單位
 - 提供正確的圖示與URL
 - 資料元素的設計與使用，必須達成資源的描述、發掘與利用

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
- Registry and template的開發（MAAT）
 - 作為計畫相關文件的上載與保存，利於評鑑及未來再利用與加值應用的依據
http://metadata.teldap.tw/design/mapping/service_apply_mapping.php

配套措施-其他

- 匯入的必備元素與值（拓展臺灣數位典藏計畫）
 - 共包括Title、Publisher、Subject、Format、Right及原始資料庫連結(Id)等六個元素

致謝

1. 拓展台灣數位典藏計畫、TELDAP聯合目錄
2. 陳亞寧、鍾豐謙、沈漢聰、城菁汝、陳慧嫻

問題與討論

鍾豐謙

fcchung@sinica.edu.tw

(02)27839766 ext. 226