

聯合目錄知識化工程

計畫主持人：陳克健 中央研究院資訊科學研究所
報告人：溫淳雅 中央研究院資訊科學研究所

2008.9.24

簡報大綱

- 一. 聯合目錄
- 二. 知識化工程
- 三. 融入未來永續經營規劃

一、聯合目錄

■ 聯合目錄是什麼？

- 為本國家型計畫所建置的目錄性展示平台
- 提供單一入口的整合性檢索功能
- 以內容主題、導覽分類等方式為藏品提供關連
- 達到資訊共享的目的

■ Metadata：典藏與學習資源的描述訊息，提供連結的訊息來源

數位典藏與數位學習 成果入口網
digitalarchives.tw

珍藏特展 目錄導覽 技術體驗 部落格 回到首頁 ENGLISH

快速導覽 主題分類 典藏機構 進階搜尋

▶ 首頁 ▶ 主題分類

搜尋 搜尋

主題小組列表

<p>生物主題</p> <p>動物界 植物界 真菌界 文獻 原生生物界</p> <p>共有 261,883 筆</p>	<p>考古主題</p> <p>考古地區 考古學文化 考古遺址 考古遺物 考古遺跡 漢代簡牘</p> <p>共有 150,669 筆</p>	<p>地質主題</p> <p>古生物 地質鑽探岩心 岩石 礦物</p> <p>共有 6,148 筆</p>
--	--	--

數位典藏與數位學習 成果入口網
digitalarchives.tw

珍藏特展 目錄導覽 技術體驗 部落格 回到首頁 ENGLISH

快速導覽 主題分類 典藏機構 進階搜尋

▶ 首頁

搜尋 搜尋

推薦特藏

超自然力的應許——消災祈福

宗教是人類最深的心靈觸動，也是反映人對自然與環境的一種和諧態度；儀式是歷經長久的生活體驗，也是宗教信仰體系下所累積對超自然的一種共同文化。不同族群對生活體驗及宗教信仰的不同，各自創造或衍生出個別的文化...

人類學主題小組



快速導覽

近代史主題

國民革命北伐 國際共產主義 孫中山 蔣中正 中國共產黨 張作霖 李大釗 四一二事件/清黨

臺灣棒球主題

張誌家 陳鎮基 林恩宇 陳金鋒 林智勝 林華韋 陳大豐 奧運棒球賽 郭源治 陳義信 紅襪少棒隊 金龍少棒隊 威爾波特少棒賽 亞洲棒球錦標賽 世界棒球錦標賽 亞洲運動會-棒球 王貞治 呂明賜 郭泰源 李居明 世界棒球二冠王 胡金龍 林英傑 王傳家 許銘傑 謝長亨 潘威倫 葉君璋 張泰山 徐生明 林易增

台灣原住民服飾主題

特色藏品



台灣棒球101

棒球是台灣最具標性的運動。1906年台灣總督府國語學校

聯合目錄內容主題分類

聯合目錄首頁

二、知識化工程

工作方向

- 建立物件metadata模型
- 建立聯結關係
 - 設計自動化工具
 - 提供典藏與學習內容自動連結工具
- 建構Teldap知識架構(Ontology)與索引典(Thesaurus)
 - Art & Architecture Thesaurus by Getty

1. 建立物件metadata模型

- 典藏品
 - 數位典藏metadata結構範例
- 網站
 - 領域
 - 產生者
 - 適用對象
 - 用途
 - 軟體形式
 - 選用數位典藏metadata規範

2. 建立典藏與學習內容聯結關係

- 主題連結：依據不同主題編寫文本
 - 標記文本中關鍵詞
 - 提供關鍵詞超連結
 - 提供相關詞超連結

文本

台灣四面環海，受到大陸沿岸南向寒流、黑潮暖流及季風漂流等作用之影響，基礎生產力豐富，極適合魚蝦貝介類之棲息繁殖，因而形成了一種佳的天然漁場，得天獨厚的地理環境，孕育了豐富而多樣化的海洋生命力。台灣四面環海，受到大陸沿岸南向寒流、黑潮暖流……

OID	Title	基本語彙	主題詞	關鍵字	自由主題詞	作者	年代	單位	URL
101	蝦類	Decapoda pandanus	蝦類	十肢目 十足目	十肢動物界 節肢動物門 十足目 十足目	呂學海	2006/01	國立自然科學博物館	#oikaaha.ndipora.tw/oid/101
102	蝦類	Nebria femineus	蝦類	十足目 十足目	十肢動物界 節肢動物門 十足目 十足目	呂學海	2006/01	國立自然科學博物館	#oikaaha.ndipora.tw/oid/102
103	大甲殼類	Stomatopoda fasciata	大甲殼類	十足目	十肢動物界 節肢動物門 十足目 十足目	呂學海	2006/01	國立自然科學博物館	#oikaaha.ndipora.tw/oid/103
104	蝦類	Rhynchops nispot	蝦類	十足目 十足目	十肢動物界 節肢動物門 十足目 十足目	呂學海	2006/01	國立自然科學博物館	#oikaaha.ndipora.tw/oid/104
105	魚類	Apogon fasciatus	魚類	魚類	魚類	呂學海	2006/01	國立自然科學博物館	#oikaaha.ndipora.tw/oid/105

2.1 開發關鍵詞超連結工具

- 典藏與學習關鍵詞辭典
- 困難與解決方式
- 斷詞技術與詞頻統計

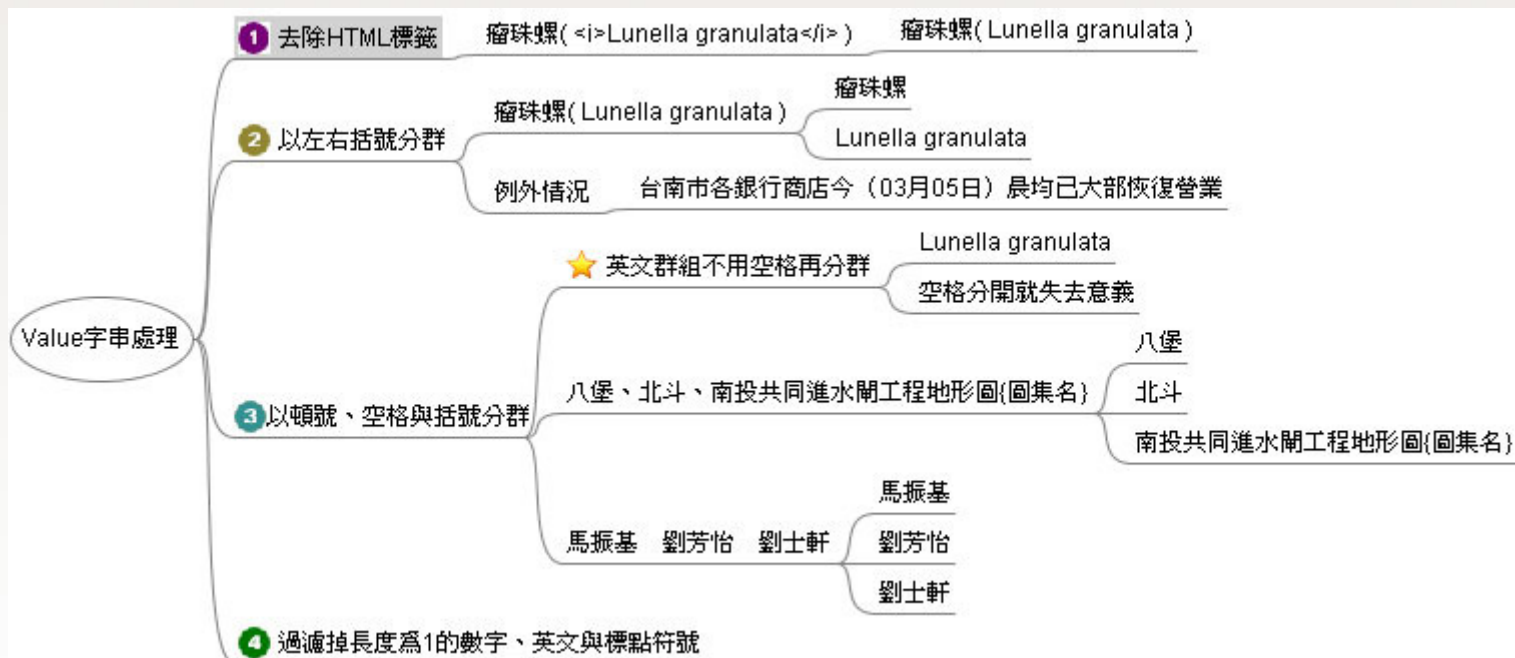
2.1.1 典藏與學習關鍵詞辭典

- 從聯合目錄彙整可用資訊，建立資料庫關聯參照
- 步驟為：
 - 從XML中抽取資料
 - 抽取資料包括：主題名、英文題名、關鍵詞、分類架構、作者、地點等
 - 配合斷詞系統，擴充專有名詞

OID	TITLE	TITLE_MAINKEYWORD	TITLE_SUBKEYWORD	CLASS	AUTHOR	ERA	LOCATION	ENGLISH_TITLE
22652	藍身大斑石斑魚...	藍身大斑	石斑魚	生物:動物界:脊索動物門...	鑑定者:鄭明...	採集日:2001...	採集地:屏東'東港...	Epinephelus tukula...
22670	斑點圓鱗'E...	斑點	圓鱗	生物:動物界:脊索動物門...	鑑定者:陳正...	採集日:199E...	採集地:宜蘭'大溪...	Scorpaena onaria ...
22672	奧利亞口孵非鯽...	奧利亞	口孵、非鯽	生物:動物界:脊索動物門...	鑑定者:鄭明...	採集日:2001...	採集地:新竹'竹北...	Oreochromis aureu ...
22675	莫三鼻口孵魚...	莫三	鼻、口孵、魚	生物:動物界:脊索動物門...	鑑定者:鄭明...	採集日:2001...	採集地:新竹'竹北...	Oreochromis mossi...
22707	烏伊蘭擬金眼鯛...	擬金眼鯛	烏伊蘭	生物:動物界:脊索動物門...	鑑定者:吳宗...	採集日:2002...	採集地:雲林'四湖...	Pempheris oualens ...

2.1.2 困難與解決方式

- 不同標準的資料整合
- 資料正規化流程
- 字串處理



2.1.3 斷詞技術與詞頻統計

- 解決分詞歧義及辨識未知詞
- CKIP word segmentation system:
<http://ckipsvr.iis.sinica.edu.tw/>
- 詞頻比較

取得候選關鍵字

NDAP:		Corpus:	
Min: 1	<input type="text"/>	Min: 0	<input type="text"/>
Max: 102366	<input type="text"/>	Max: 581857	<input type="text"/>
<input type="text"/> (1 ~ 19087)		<input type="text"/> (0 ~ 8082)	

	<input type="text"/> 候選關鍵字 ↓	斷詞結果	NDAP	Corpus	詞性	Nb	來源長度	Oid
1	义干蘇蘇卡倫		1	0	Nb		6~	770306
2	义哈爾		1	0	Nb		6~	3130644
3	万里加		8	0	Nb		6~	還多...
4	仇		1	0	Nb		6~	1557575
5	仇氏		1	0	Nb		6~	2989683
6	无垢淨光佛		2	0	Nb		6~	1807341 、 1811846
7	无畏		5	0	Nb		ALL	還多...
8	全町三		2	0	Nb		6~	3061736 、 3061737
9	全僅博		4	0	Nb		6~	還多...
10	全仁		1	0	Nb		6~	602061
11	全木交		1	0	Nb		6~	1575613
12	全立		232	0	Nb		ALL	還多...
13	全立分田		1	0	Nb		6~	1783917

2.2 關鍵詞超連結雛形範例

聯合目錄管理系統

Step1: 請選擇其中一種資料來源

- A. 上傳所需轉換之檔案
- B. 輸入所需轉換HTML文本

Step2:

2.2 關鍵詞超連結雛形範例

■ 取得相關關鍵字(可勾選)

聯合目錄管理系統

[新聞](#) [快速導覽](#) [特色藏品](#) [時間檢索](#) [流量報表](#) [資源統計](#) [檢索詞統計](#) [藏品異動](#) [關鍵字連結轉換](#)

Step1: 輸入原始HTML資料

```
<h3>水產試驗所魚蝦貝類標本數位典藏</h3>
<p class="ps"> 臺灣沿近海海洋生物標本之數位典藏—魚、貝、甲殼類</p>



<p>台灣四面環海，受到大陸沿岸南向寒流、黑潮暖流及季風漂流等作用之影響，基礎生產力豐富，極適合魚蝦貝介類之
棲息繁殖，因而形成了一極佳的天然漁場，得天獨厚的地理環境，孕育了豐富而多樣化的海洋生命力。
<br><br>
農委會水產試驗所為我國唯一的水產試驗機關，數十年來對於台灣周邊海域的水產生物資源調查與魚蝦貝類人工繁殖技術的研究不遺餘力。從
1929年成立以來，典藏了許多寶貴的水產生物標本，成為台灣海洋研究與漁業發展的重要資產。典藏的標本經研究人員有系統的整理、清查、重新
鑑定、資料登錄和標籤製作處理，已逐步建立魚類標本資料庫。同時，將每一標本原件影像數位化，並完成網站的建置，提供線上查詢與應用。二
年來，已建立將近2,000筆魚、貝、蝦類及頭足類的標本資料，提供快速導覽與搜尋表單的搜尋功能，可快速搜尋您需求的水產生物資訊。若您想進
一步瞭解相關的種類資訊，請上網查詢，網址如下：
```

Step2:

Step3: 勾選關鍵字清單:

欄位顯示: [1欄](#) [2欄](#) [3欄](#) [4欄](#) [5欄](#) [6欄](#) [7欄](#)

關鍵字長度: [1個字以上](#) [2個字以上](#) [3個字以上](#) [4個字以上](#) [5個字以上](#)

<input checked="" type="checkbox"/> 全部關鍵字	<input checked="" type="checkbox"/> 紅皮刀	<input checked="" type="checkbox"/> 印度洋	<input checked="" type="checkbox"/> 康吉鰻	<input checked="" type="checkbox"/> 太平洋
<input checked="" type="checkbox"/> 馬拉巴	<input checked="" type="checkbox"/> 巨口魚	<input checked="" type="checkbox"/> 成功	<input checked="" type="checkbox"/> 旗魚	<input checked="" type="checkbox"/> 人員
<input checked="" type="checkbox"/> 格氏光	<input checked="" type="checkbox"/> 日本	<input checked="" type="checkbox"/> 台灣	<input checked="" type="checkbox"/> 深海	<input checked="" type="checkbox"/> 條紋
<input checked="" type="checkbox"/> 臺灣	<input checked="" type="checkbox"/> 草魚	<input checked="" type="checkbox"/> 口魚	<input checked="" type="checkbox"/> 斧魚	<input checked="" type="checkbox"/> 海洋
<input checked="" type="checkbox"/> 印度	<input checked="" type="checkbox"/> 光魚	<input checked="" type="checkbox"/> 黑潮	<input checked="" type="checkbox"/> 烏魚	<input checked="" type="checkbox"/> 皮刀
<input checked="" type="checkbox"/> 銀斧	<input checked="" type="checkbox"/> 鱸仔	<input checked="" type="checkbox"/> 紅鱸	<input checked="" type="checkbox"/> 石斑	<input checked="" type="checkbox"/> 珠鱸
<input checked="" type="checkbox"/> 香魚	<input checked="" type="checkbox"/> 拉巴	<input checked="" type="checkbox"/> 直線	<input checked="" type="checkbox"/> 長身	<input checked="" type="checkbox"/> 笛鯛
<input checked="" type="checkbox"/> 海水	<input checked="" type="checkbox"/> 黃鰭	<input checked="" type="checkbox"/> 龍尖	<input checked="" type="checkbox"/> 黑斑	<input checked="" type="checkbox"/> 閃電
<input checked="" type="checkbox"/> 石鱸	<input checked="" type="checkbox"/> 雨傘	<input checked="" type="checkbox"/> 格氏	<input checked="" type="checkbox"/> 大小	<input checked="" type="checkbox"/> 處理
<input checked="" type="checkbox"/> 鳥嶼				
<input checked="" type="checkbox"/> 應用				

Step4:

數位典藏網路核心平台計畫

2.2 關鍵詞超連結雛形範例

■ 產生轉換後的文本

[> 首頁](#) > [> 特色藏品](#) > [> 檢視網頁](#)

特色藏品內容

水產試驗所魚蝦貝類標本數位典藏

臺灣沿近海海洋生物標本之數位典藏—魚、貝、甲殼類

典藏標本的特色是以海洋性及深海魚種為主，標本採集的範圍遍及台灣周邊海域及各大小島嶼、南北太平洋及印度洋公海區，甚至遠達南極海域，魚種的分布水深由表層至4,000餘公尺，除了常見的種類外，也有許多罕見或瀕臨絕種的魚類，像是罕見的棘鰻、日本短尾康吉鰻、後鰭魚、格氏光巨口魚、長身櫛口魚、閃電燭光魚、長銀斧魚等。深海魚類最大特徵是眼睛退化、牙齒特長或具有發光器等，因種類而異。牙齒長能讓好不容易才捕到的獵物不易脫落；發光器則具有誘捕生物、吸引異性、辨識同類等功能。發光器在體上的分布位置與排列形式，也是深海魚種類鑑定的重要依據。例如採集自2,050公尺深的長銀斧魚(圖1)，身上具有明顯的發光器，由胸部至臀部幾乎成一直線排列，尾部發光器之間距較大。

人工繁養殖成功的魚種，是本所典藏標本的另一特色，包括1960年代的草魚、鱧魚、烏魚，1970年代的香魚、虱目魚、黑鯛，1980年代的馬拉巴石斑(珠鱸仔)、黃鰭鯛(赤鰭仔)、龍尖，1990年代的花軟唇(石鱸)、黑斑紅魷(紅鼓魚)、黃姑魚等種類。

此外，水族館有淡水魚類如固魚、石賓、青魚、鯽魚、銀鱸、條紋鱸、虹鱔、西伯利亞鱒等，海水魚類有虱目魚、黑鯛、馬拉巴石斑、鞍帶石斑、川紋笛鯛、黃鰭鯛、銀紋笛鯛、紅鼓魚、海鱚等活生生的經濟食用魚類。水產陳列室展示多種大型魚類乾製標本，如雨傘旗魚、月魚(紅皮刀)、翻車魚等，有海豚、齒鯨及鬚鯨實體標本，還有罕見的深海菱鰭魷等，種類繁多，歡迎參觀。

行政院農委會水試所海洋漁業組 撰寫

National Digital
Archives Program
Taiwan

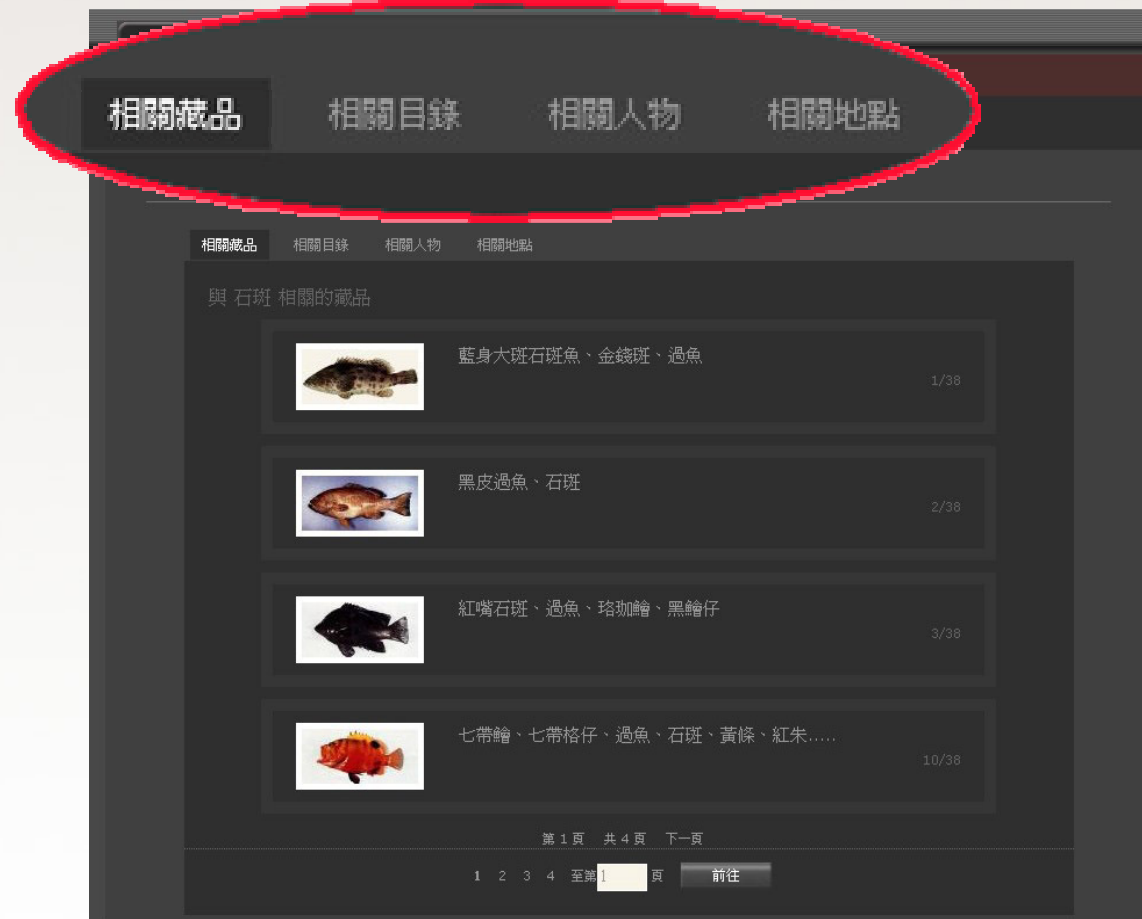


[關於聯合目錄](#) [常見問題](#) [匯入介紹](#) [聯絡我們](#)

數位典藏國家型科技計畫 聯合目錄系統建置計畫
Union Catalogs Program, National Digital Archives Program TAIWAN.

2.2 關鍵詞超連結雛形範例

- 提供關鍵詞的相關內容
 - 轉換後的文本，其關鍵詞可超連結至相關內容



3. 建構知識架構(Ontology)與索引典(Thesaurus)

■ 分類關係

■ 同義關係

□ 蘇軾=蘇東坡= Sushi

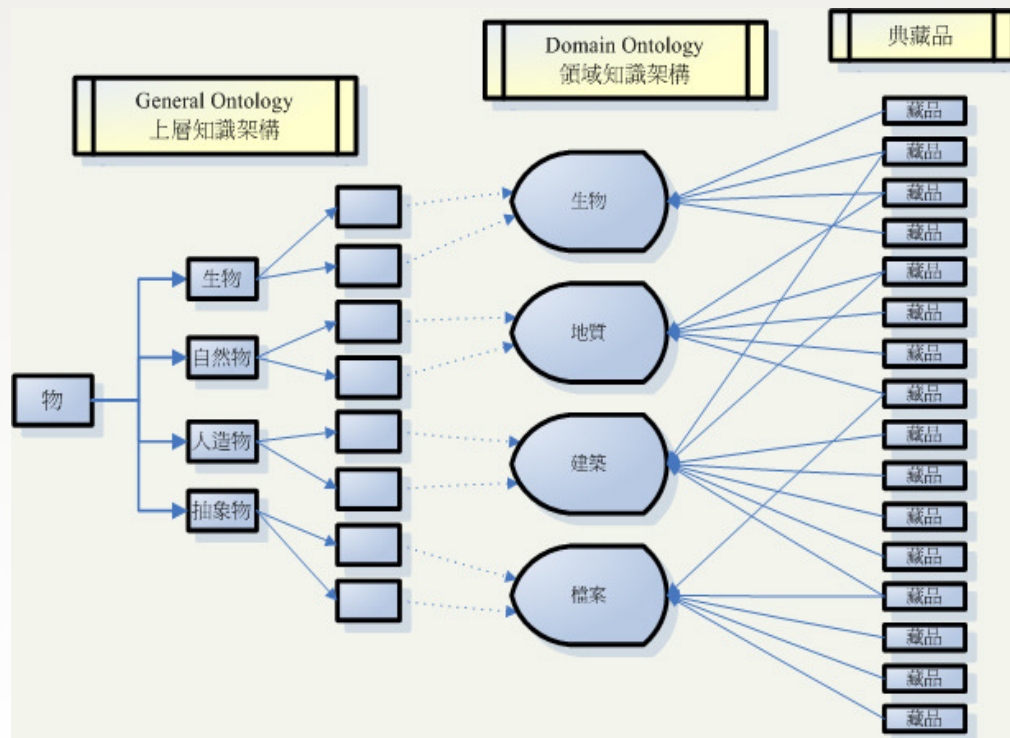
□ 鄭成功=延平郡王

■ 反義關係

■ 上下位關係

□ 器物→[陶器、瓷器]/[杯、盤、碗、甕]

■ 領域、時間、地理、事件、因果等關係



三、融入計畫辦公室永續經營規劃

- 技術發展
- 內容合作

■ 技術發展

- 建立時間、地點關聯—時空座標
- 建立中英索引典—Getty AAT
- 建立關鍵詞與關鍵詞關係知識架構—內容主題知識架構
- 設計、擴展關鍵詞超連結標記工具及延伸工具
- 標記機構典藏間之超連結
- 中文分析—分析文本內容
- 連結排序—比對典藏物件以進行相關性排序
- 建置典藏與學習資源內容知識架構管理檢索平台
- 從聯合目錄構建藏品、關鍵詞等資料庫
- 工具及資料更新、環境維護及推廣

■ 內容合作

□ 計畫內合作：

- ◆ 規劃物件的表達模型(metadata)
- ◆ 所有物件填具其metadata
- ◆ 提供工具供編寫不同主題之知識內容，建立典藏與學習資源內容之超聯結
- ◆ 充實所選主題典藏與學習內容資源
- ◆ 建構國中小學教科書與典藏與學習資源間的超聯結

□ 尋求向外擴展合作對象：例如 維基百科

報告完畢
敬請指教
