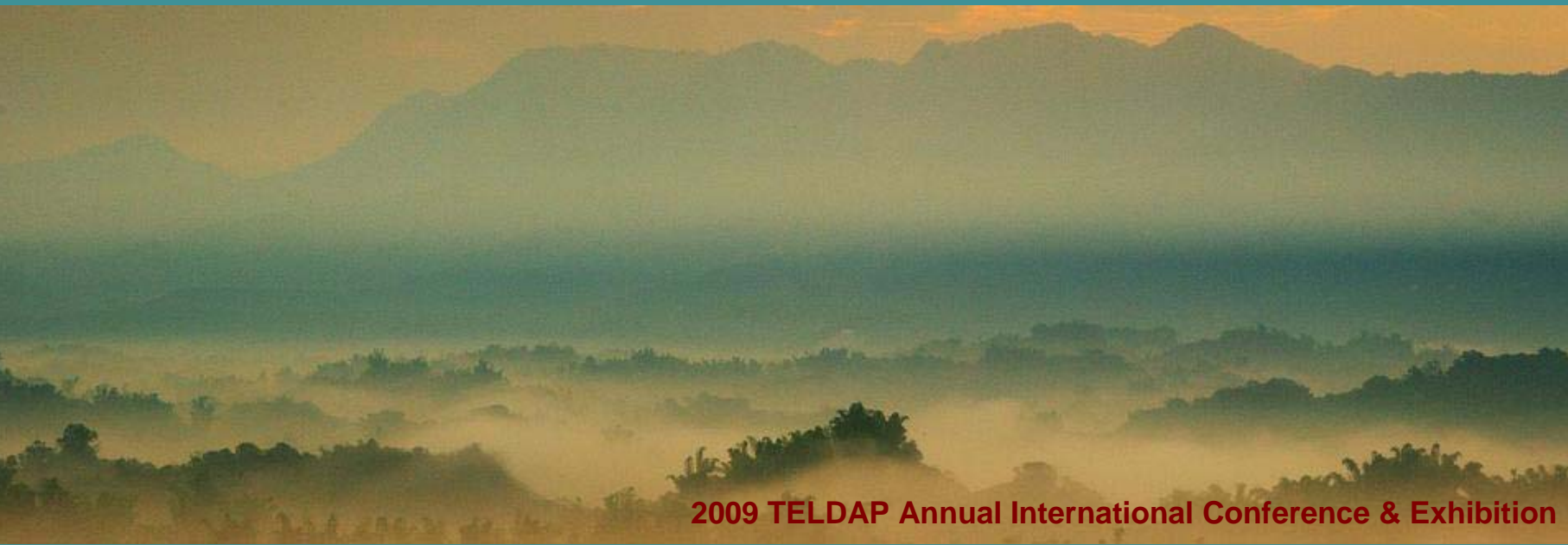


**Hsueh-Hua CHEN & Sophy Shu-Jiun CHEN**

National Taiwan University, Taiwan | Academia Sinica, Taiwan



**2009 TELDAP Annual International Conference & Exhibition**

# 知識組織系統的多語言等同對照分析

**Analysis of Multilingual Equivalence Mapping for Knowledge Organization Systems**

# Background

1

數位圖書館的重要特色之一，是可以跨越時間與空間，讓全球皆有機會近用(**access**)數位資源。

One of the key features of digital libraries is to provide **global access to digital resources** beyond the boundaries of time and space.

2

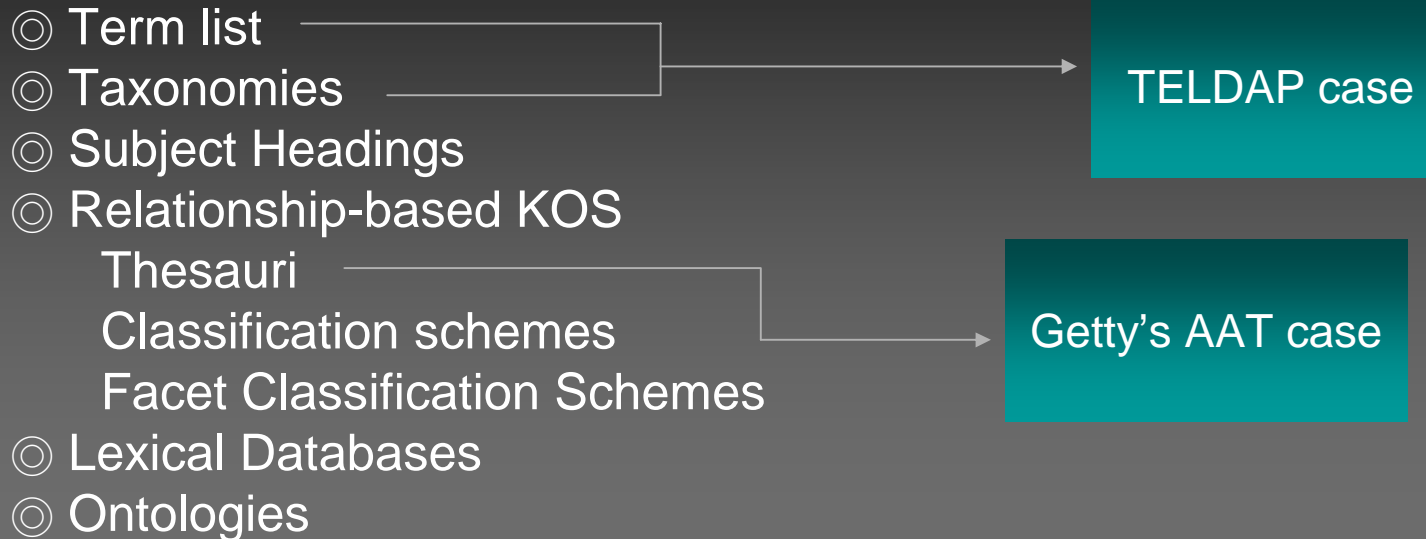
但是，如何跨越語言的障礙，讓不同語言的使用者可以檢索到典藏在世界各地的數位內容，是目前數位圖書館服務面臨的重要研究課題。

One of the most important tasks of digital libraries is to overcome **language barrier** and allow **multilingual access** to material worldwide

# Background 3

This study aims to explore the issue of multilingual access to digital libraries from knowledge organization system (KOS) perspective.

Hodges (2000) & JISC (2006) provide typologies of KOS according to structure. 4



# Chinese AAT Project Objectives

增加AAT中文等同關係詞彙（目前已有荷蘭文、法文、西班牙文、義大利文；德文翻譯計畫正在進行中），除此，也增加從TELDAP觀點所分析出來的新詞彙至AAT

Incorporation of Chinese terms into AAT (there are Dutch, French, Spanish, Italian; and German translation currently Underway), and **adding new candidate terms** into AAT from perspectives of TELDAP

建立TELDAP多語言索引典的基礎架構

Establish the basic infrastructure for the TELDAP multilingual thesaurus .

數位圖書館的多語檢索系統的基礎研究

Fundamental research on multilingual data retrieval of digital libraries


# Chinese AAT Project

## Specific Research Questions



藝術與物質文化領域，不同語言及不同知識組織系統的詞彙之等同關係的匹配類型為何？

**What is the range of terminology equivalence between...?  
Different languages (Chinese and English) and  
KOSs (TELDAP and AAT), such as art and  
material culture**



承上，不同詞彙系統（來源詞彙與目標詞彙）的概念結構之異同性類型為何？

**What are the similar types of conceptual structures  
between TELDAP and AAT?**

# Methodology

## Selection of Terms

以「故宮器物後設資料需求規格書」選擇其中**60**個控制詞彙，進行與**AAT**詞彙等同關係對照實作。此**60**個控制詞彙，含括**AAT**層面架構中的**6**個層面，分別是「物理屬性」、「風格與時期」、「代理者」、「活動」、「材質」及「物件」層面；只有「相關概念」層面尚未包括在本次先導研究範圍。

**Select 60 controlled vocabularies from the Metadata Specification of National Palace Museum (Department of Antiquities) for AAT equivalence mapping purpose, all facets included, except for Associated Concepts facet.**

# Example of Selected Terms

Selected terms of TELDAP	AAT FACET and Hierarchy
全器形制－植物－蔬果式－葫蘆形 全器形制－幾何－平面－三角形 局部形制－頸－長頸	PHYSICAL ATTRIBUTES FACET Attributed and Properties Hierarchy
紋飾－幾何－雷紋－雲雷紋 紋飾－幾何－點狀紋－乳丁紋	PHYSICAL ATTRIBUTES FACET Design Elements Hierarchy motifs
窯系－明代官窯系 窯系－景德鎮窯系 窯系－定窯系	STYLES AND PERIODS FACET Styles and Periods Hierarchy <Chinese ceramics styles>
考古學文化－仰韶文化 考古學文化－良渚文化	STYLES AND PERIODS FACET Styles and Periods Hierarchy <Chinese Neolithic periods>
紋飾－人物－宮廷人物－后妃 紋飾－人物－佛道人物－彌勒佛	AGENTS FACET People Hierarchy religious (people)

# Methodology

## Data Collection

### Three types of data collection

- ◎ **Mapping analysis logs from researchers and participants**  
(60個詞彙，143份) 研究者與實驗者的實作記錄
- ◎ **Mapping analysis workshops**  
研究者與實驗者的工作坊記錄 (5場)
- ◎ **Email communications with maintainers of the term schemes**  
詞彙來源單位的溝通記錄

A total of 12 people, including 11 participants from TELDAP and Related programs, contribute in the data collection process



# Data Analysis

在範圍上，本研究的分析工作可分兩個部分

- ◎ 第一部分目的是要識別TELDAP與AAT詞彙之間的等同關係之匹配類型。主要工作包括：
  - 確認TELDAP英譯詞彙
  - 判斷與確認此詞彙在AAT可能的等同概念之詞彙 (concept-based)
  - 賦予等同關係的匹配類型
- ◎ 第二部分，是觀察與判斷二個來源詞彙在TELDAP與AAT概念結構的異同性。

The data analysis can be divided into two parts

- ◎ Part 1, identify types of possible matches of terms between TELDAP and AAT, which includes :
  - Confirm English translation of the term in TELDAP
  - Identify types of possible matches (concept-based)
  - Confirm types of equivalence matches
- ◎ Part 2, observation and identification of similar types of conceptual structures between TELDAP and AAT

# Preliminary Findings

## Frequencies of equivalence mapping types

「完全等同關係」（35個，58%）是本研究結果出現最高頻率的匹配類型，其次依序是「部份等同關係（種－屬附屬關係）」（18個，30%）、「不等同關係」（6個，10%）、「不完全等同關係」（1個，2%）。

Exact Equivalence (35 terms, 58%) has the highest match ratio, followed by Partial Equivalence (18 terms, 30%), Non-equivalence (6 terms, 10%), Inexact equivalence (1 term, 2%)

# Preliminary Findings

## Range of Equivalence

Match Type code	Definition	Sample	
		TELDAP's examples	AAT's examples
1-1	Exact Equivalence	仰韶文化	<b>yangshao</b>
1-2	Exact cross-reference match	掐絲	<b>filigree enameling</b>
2-1	Inexact Equivalence	香爐	<b>censers</b>
3-2	Partial Equivalence (species-genus relationship ...Subordination)	乳丁紋	<b>dots</b>
5-1	Non-equivalence (culture-dependent)	龍泉窯	X
5-3	Non-equivalence (beyond scope)	彌勒佛	X

# 仰韶文化 = Yangshao

Match Type code	definition	Examples	
		TELDAP term	AAT term
1-1	Exact Equivalence	仰韶文化(<考古學文化)	Yangshao (<Chinese Neolithic periods>, <Chinese prehistoric periods>, ... Styles and Periods)

仰韶文化



仰韶文化 玉盤  
5000 B.C.-3000 B.C.

# 掐絲 = filigree enameling

Match Type code	Definition	examples	
		TELDAP term	AAT term
1-2	Exact cross-reference match	掐絲(<技法) (參見 掐絲琺瑯)	filigree enameling (cloisonné, enameling, ... Processes and Techniques)



清 乾隆 掐絲琺瑯鳧壺  
1736 A.D.-1795 A.D.

Exact Equivalence  
 掐絲琺瑯 ←→ filigree enameling  
 用(USE) 掐絲 (preferred term)

# 香爐 +/- censers

Match Type code	Definition	Examples	
		TELDAP term	AAT term
2-1	Inexact Equivalence	香爐	<b>Censers</b> (<ceremonial containers>, <containers by function or context>, ... Furnishings and Equipment)



Note: Refers to containers **with perforated covers** used for burning incense in a ritual context, **especially ecclesiastical**; usually of metal or ceramic.

ID: 300198814



# 乳丁紋 < dots

Match Type code	Definition	Examples	
		TELDAP term	AAT term
3-2	部分等同關係（種－屬附屬關係） <b>Partial Equivalence</b> (species-genus relationship ...Subordination)	乳丁紋(<點狀紋<幾何<紋飾)	<b>dots</b> (<geometric motifs>, motifs, ... Design Elements) <b>ID: 300010145</b>



商後期 鈎連乳丁紋羊首罍  
**Ram's-head Lei with Hooks and Dots**  
Late Yin period (1300-1046 BC)  
Height: 37.3 cm, rim diameter: 31.3 cm

# 龍泉窯 X (AAT term)

Match Type code	Definition	Examples	
		TELDAP term	AAT term
5-1	不等同關係（文化獨特性）  <b>Non-equivalence (culture-dependent)</b>	龍泉窯(<窯系) Longquan Kilns	×



- Styles and Periods Facet
- .... Styles and Periods
- ..... <styles and periods by region>
- ..... Asian
- ..... East Asian
- ..... Chinese
- ..... <Chinese styles and periods>
- ..... <Chinese styles (guide term)>
- ..... **<Chinese ceramics styles>**
- ..... Ru



# 彌勒佛 X (AAT term)

Match Type code	Definition	Examples	
		TELDAP term	AAT term
5-3	不等同關係（超越收錄範圍）  Non-equivalence (beyond scope)	彌勒佛 (<佛道人物<人物<紋飾) Maitreya Buddha	X

作者不詳（清）。[中文品名：彌勒佛（72-00407）]。《數位典藏聯合目錄》。

<http://catalog.ndap.org.tw/?URN=1197865>（2009/02/07瀏覽）。



# Preliminary Findings

## Similarity Types of Conceptual Structures

根據本研究初步發現，不同系統的概念結構相似性範圍可依程度暫時

分為4種類型，依序是：

**Type 1** 架構相似(**Similar Structure-modeling**)：可採用模仿方法 (**modeling**) 將目標詞彙 (**AAT**) 的部份架構移植到來源詞彙 (**TELDAP**) 的系統

**Type 2** 架構相似(**Similar Structure-expanding**)：但需再擴充或修訂目標詞彙 (**AAT**) 的架構

**Type 3** 架構不相似(**Dissimilar**)：來源詞彙可部份等同對照至目標詞彙

**Type 4** 架構缺乏(**Lack of Structure**)：來源詞彙 (**TELDAP**) 無法等同對照至目標詞彙 (**AAT**)

**According to the preliminary findings, the similarity of the conceptual structures of different systems can be categorized into 4 types**

# Preliminary Findings

## Type 1 架構相似 Similar Structure

本研究對於「架構相似」的定義，係指來源知識組織系統（TELDAP）的階層邏輯與目標知識組織系統（AAT）相同，且與TELDAP的脈絡結構不衝突、可相容的情況。通常AAT的概念結構比TELDAP具有更多，或更寬廣的層級，因此可以採用模仿方法（modeling）將AAT的該部份概念結構移植到TELDAP。

**Similar Structure, meaning the hierarchical logic of the KOS (TELDAP) is the same as the target KOS (AAT), with no structural Conflict. The conceptual structure of AAT is usually more defined Than TELDAP. By utilizing the method of modeling, partial AAT Conceptual structure can be transferred to TELDAP.**

# Preliminary Findings

Type 1 架構相似  
Similar Structure



TELDAP

people (agents) > monarchs >  
Empresses

AAT

Agents Facet > people > **people**  
**(agents)** > <people by  
occupation> >  
<people in government and  
administration> > rulers (people)  
> **monarchs** > **empresses**



# Preliminary Findings

## Type 1 架構相似 Similar Structure



代理者層面 > 人物(層級) > 人物 > (人物，依據職務區分) > (公部門的人物) > 統治者 > 宮廷人物 > 后妃

Agents Facet > people > people (agents) > <people by occupation> > <people in government and administration> > rulers (people) > monarchs > empresses



Agents Facet代理者層面

.... **People**人物層級

..... people (agents) 人物

..... <people by occupation> <人物，依據職務區分>

..... <people in government and administration> <政府與管理的人物>

..... rulers (people) 統治者

..... **monarchs** 宮廷人物

..... **empresses** 后妃

# Preliminary Findings

## Type 2 架構相似 Similar Structure

架構相似—但需再擴充或修訂目標詞彙的架構

**Similar Structure, but required further expansion or correction of the structure of target language.**

TELDAP詞彙代表的概念常出現中國文化的獨特性，因此需在AAT現有的概念結構中，再更深入往下層延伸細部結構、或往寬廣的層級新增整套概念結構。

**The concepts of the terms in TELDAP usually convey a uniqueness only known in Chinese culture, therefore we need to develop a more intricate structure.**

具中國文化獨特性或表達性概念的詞彙，包括：「窯系」、「考古文化」、「全器形制」、「書體」及「紋飾」等概念結構。

**These unique concepts include :窯系(kiln), 考古文化, 全器形制 書體(script) ,紋飾(motifs)**

# Preliminary Findings

Type 2 架構相似  
Similar Structure

以「書體」(script) 為例：



clerical script (隸書, lishu)



seal script (篆書)



semi-cursive script (行草書)



regular script (楷書, Kaishu)



bird-and-insect script (鳥蟲書)

# Preliminary Findings

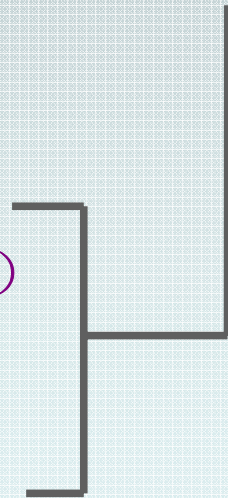
## Type 2 架構相似 Similar Structure

### .... Objects Facet

- ..... Components (Hierarchy Name)
- ..... components (objects)
- ..... <components by specific context>
- ..... <information form components>
- ..... <script and type forms>
- ..... **scripts (writing)**
- ..... <scripts by form>
- ..... **<Arabic scripts>** 阿拉伯書體
- ..... **<Chinese scripts>** (中文書體)
- ..... **bird-and-insect script** (鳥蟲書)
- ..... **clerical script** (隸書, lishu)
- ..... **regular script** (楷書, kaishu)
- ..... **seal script** (篆書)
- ..... **semi-cursive script** (行草書)

### **Suggest:**

Create a set of new  
guide term and  
candidate terms





# Preliminary Findings

## Type 3 架構不相似 Dissimilar Structure

所謂「架構不相似」，在本研究是則表示TELDAP詞彙呈現的概念結構與AAT表現的概念結構之分類邏輯不相同，但TELDAP詞彙可以完全等同或部分等同地對照到AAT詞彙。如：釉色

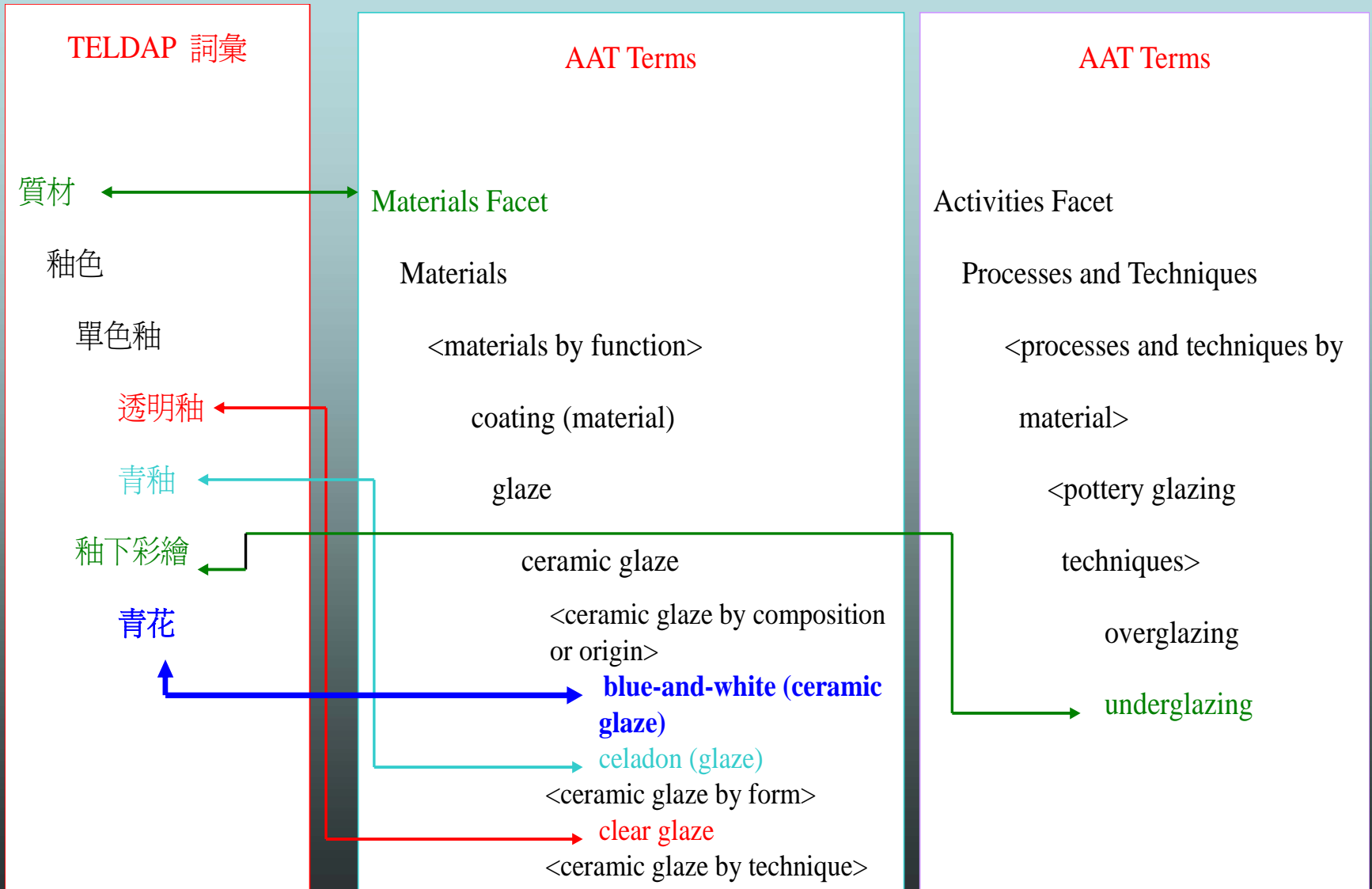
**Dissimilar Structure, the conceptual structure of the term in TELDAP is different from AAT, but the term can either be exactly matched or partially matched to the term in AAT, for example: ceramic glaze**

## Type 4 架構缺乏 Lack of Structure

「架構缺乏」，則是指TELDAP詞彙呈現的概念結構無法在AAT發現相似結構，相關詞彙也無法於AAT對照到任何等同關係的詞彙。如：局部形制

**Lack of Structure (no structure), neither similar conceptual structure nor related term can be found in AAT,**

# ceramics glaze



# Closing Remarks

相同領域的不同語言、不同知識組織系統之間的詞彙，具有高度相容性（包括詞彙等同關係程度，及概念結構相似性高），雖然彼此間會有文化獨特性產生的概念與詞彙，也許是互相尚未收入的。

**the terms in different languages and KOSs under the same discipline are highly compatible (including equivalence mapping types, similarity between the conceptual structures). It is likely that unique cultural-related concepts and terms are not yet included.**

本研究個案的來源語言與目標語言之知識組織系統，皆各自保留未來修改的彈性

**The KOSs of the source language and the target language retain their flexibility for future revisions**

未來可進一步研究，如何發展與設計，可以容納更多視野與觀點的概念結構

**The next step is to develop and design a conceptual structure to integrate a wider variety of ideas**

本研究未來將加入更多的學科專家，以更詳盡而精確地檢視這些詞彙等同關係與概念結構的適當性與準確性，進而提升本研究結果的品質

**This research will invite more subject experts to closely examine the accuracy of the equivalence mappings and conceptual structures of the terms for quality purposes.**

# Closing Remarks

- 本研究對於TELDAP與Getty's AAT合作案，有非常直接的助益。建議可以針對「完全等同關係」的詞彙，進行雙語的建置與互相連結。而對於「部分等同關係」的詞彙，若是被進一步確認是重要的概念詞彙，則可以產生新詞彙，並貢獻至AAT，如此經由亞洲區域觀點的加入與連結，可豐富AAT本身的完整性與多元文化性之價值。

This research is extremely beneficial for the collaboration between TELDAP and Getty's AAT. It is recommended to establish bilingual association links between exact equivalence terms. And also, after confirming the importance of the partial equivalence concepts, these terms can be contributed into AAT. By entering perspectives of the Asia region, it ultimately broadens AAT's scope on multicultural heritage.

# Closing Remarks

- 中文「藝術與建築索引典」的設計發展，經由TELDAP的個案研究，初步結果發現AAT的概念架構與詞彙似乎可以包括大部分TELDAP需求，且提供更大的脈絡架構，及連結TELDAP的深度性詞彙。這對於知識組織系統的發展與互通性，具有很大的意義，並且可以節省從頭開始建置的成本。

Regarding the development of Chinese AAT, through the case study of TELDAP, preliminary results indicate that the concept structures and terms of AAT meet nearly all the demands of TELDAP. AAT also provides larger context infrastructure, and links to in-depth terms in TELDAP. This is especially helpful for the development and interconnection between KOSs, and cost-effective as well.

# Closing Remarks

- 基於本研究初步發現「完全等同」(58%)與「部分等同」(30%)關係，佔TELDAP與AAT詞彙間的大多數比例(88%)；以及AAT本身已建立及不斷建置中的多語言(如：西班牙文、法文、義大利文、德文等)，加上本研究對於中文等同關係的研究，將預期可以精進與貢獻於未來TELDAP多語言資訊檢索的環境。
- Based on the preliminary results, matches of Exact Equivalence terms (58%) and Partial Equivalence terms(30%) between TELDAP and AAT are 88% in total. While AAT is expanding its multilingual links (such as Spanish, French, Italian, and German), with the addition of Chinese equivalence mapping from this research, it will help to perfect the multilingual data retrieval of TELDAP in the foreseeable future.

## Special Thanks to the following people for their assistance in our study

陳克健 · 城菁汝 · 林彥宏 · 王希純  
鍾豐謙 · 曾偉綾 · 龔裕晴 · 周永怡  
吳岱晏 · 林慧菁 · 王怡心 · 劉靜頻  
Patricia Harpring