Methodological Considerations for Developing the Art & Architecture Thesaurus in Chinese and its Applications

> Sophy Shu-Jiun CHEN Institute of History and Philology, Academia Sinica, Taiwan

> > NKOS Consolidated Workshop 2020 2020. 09. 10

AAT in Chinese Language

Developed by the Academia Sinica Center for Digital Cultures (ASCDC) of Taiwan

http://aat.teldap.tw/



ITWG The International Terminology Working Group



Pre-ITWG Meeting		ITWG Meeting		5.	2014.09
1. 2008.07		1.	2010.08	6.	2016.08
2.	2009.07	2.	2011.11	7.	2017.08
۷.	2009.07	3.	2012.01	8.	2019.02
		4.	2013.01	9.	2020.02

AAT as a Knowledge Tree

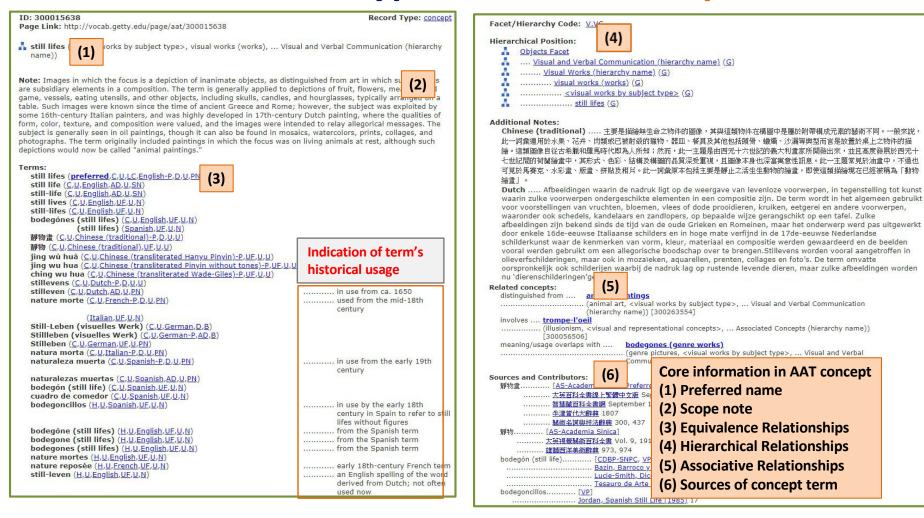
- AAT is a multilingual and hierarchical thesaurus structured in 7 facets and 21 hierarchies
- Current totals: 57,500 records (2016)
- Includes generic terms to describe art, architecture, design, artifacts, archival materials, archaeology, visual surrogates and other material culture



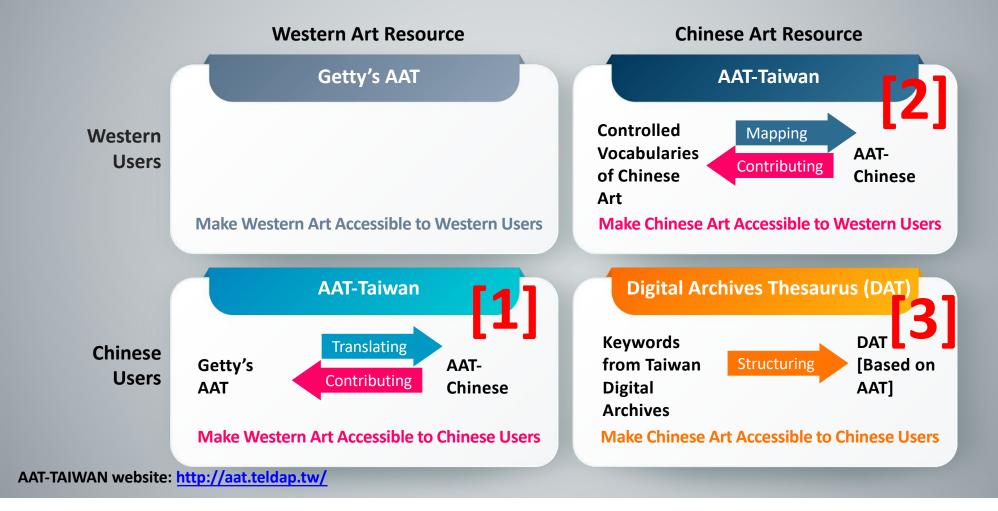


Rose Paintings and 學院藝術 Documents of Chen Zhi-chi/ academic Showa 玫瑰/陳植棋畫作與文書 1日本昭和 rose (color) 油畫(技術) oil paintings (technique) 玫瑰(1927)。選自: 數位曲藏與數位學習聯 合目錄》 https://catalog.digitalarchives.tw/it em/00/7f/86/b1.html

Record types in AAT: Concept



Chinese-Traditional Version of AAT: The Purpose



Working Model of AAT-Taiwan

		m English Chinese _{中文化}	to			From Chines English 在地化&英文(_
Term Collection					Term Selection 詞彙選擇	Literacy Warrant 文獻保證	V	Object Varrant 就品保證
Semantic Analysis					Identity Sources 誦意確認	Analysis o Qualia Structure 屬性分析		
Equivalence Mapping					Confirm Eng. Translation 英譯詞確認	Mapping 對應	Тур	dentify bes of EM 基類型分析
Research on KOS Cons	struction				Equivalence Relationship 等同關係	Hierarchical Relationship 屠級關係	Re	sociative lationship 続想關係
Expert Group					Content Verification 内容審定	Content Revision 内容修訂		
Concept Construction					Scope note Creation 新增範圍註			
Translation	Scope Note 範圍註		Term 詞彙		Scope Note 範圍註		Term 詞彙	
Proofreading	Scope Note Proofreading 範圍註校訂	Term Proofreading 詞彙校訂	Synonyms Collecting 同義詞蒐集	Three Types of Pinyin 三種拼音	Scope Note Proofreading 範圍註校訂	Term Proofreading 詞彙校訂	Synonyms Collecting 同義詞蒐集	Three Types of Pinyin 三種拼音
Disambiguation			Disambigua- tion 消歧義				Disambigua- tion 消成義	
Expert Group	Examination 審定		Revision 修正		Examination 審定	Revision 修正		

From English to Chinese

Make Western Art Accessible to Chinese Users



From Chinese to English

Make Chinese Art Accessible to Western Users



Conceptual Mapping of the NPM Terms to AAT

- The study takes a total of 2100 Chinese controlled vocabulary terms from the National Palace Museum in Taiwan, and a total of 35,314 English concepts of AAT, and performing semantic mapping between Chinese and English terms. The study asked research questions including:
- A. To what degree of semantic interoperability can be achieved between the two controlled vocabularies?
- B. What patterns of interoperability are involved in the field of Chinese art in the bilingual conceptual structure?

Example of mapping sheet between NPM and AAT terms

- 34	A	В	С	D	E	F
1	NPM id	NPM Term	AAT id	Mapped Terms in AAT	AAT Facet	AAT Hierarchy
2	P01001	書法	300266660	calligraphy (visual works) (G)	V	V.VC
3	P01002	繪畫	300033618	paintings (visual works) (G)	V	V.VC
4	P01003	緙絲	300227844	kesi (G)	М	M.MT
5	P01004	刺繡	300264024	embroidery (visual works) (G)	V	V.VC
6	P01007	刺繡	300103035	rubbings (visual works) (G)	V	V.VC
7	P02001	白描	300106458	Bai miao (G)	F	F.FL
8	P02002	工筆	300310140	gongbi (G)	К	K.KT
9	P02003	寫意	300310141	xieyi (G)	К	K.KT
10	P02005	沒骨	300310139	boneless	К	K.KT
11	P02007	指畫	300033706	finger paintings (visual works) (G)	v	V.VC
12	P02008	油畫	300033799	oil paintings (visual works) (G)	V	V.VC
13	P03012	山徑	300000626	trails (recreation areas) (G)	V	V.RG
14	P03017	瀑布	300008736	waterfalls (natural bodies of water) (G)	V	V.RD
15	P03018	靈	300343840	clouds (G)	В	B.BM
16	P03019	田疇	300008632	farmland (G)	V	V.RD
17	P03020	煙、霧	300055374	fog (G)	В	B.BM
18	P03021	石壁、峭壁	300008749	cliffs (G)	V	V.RD
19	P03022	岩洞	300008746	caves (G)	v	V.RD
20	P03023	島嶼	300008791	islands (G)	V	V.RD
21	P04001	帝王	300025480	emperors (G)	Н	H.HG
22	P04002	聖賢	300150555	saints (G)	н	H.HG
23	P04004	官員(臣)	300025457	public officers (G)	н	H.HG
24	P04007	仕女	300375129	gentlewomen (G)	Н	H.HG
25	P04008	孩童	300025945	children (people by age group) (G)	Н	H.HG
26	P04011	農夫	300025607	farmers (G)	Н	H.HG
27	P04013	樵夫	300025610	loggers (G)	Н	H.HG
28	P04016	圉人	300136457	horse breeders (G)	н	H.HG
29	P04017	商販	300025245	merchants (G)	Н	H.HG
30	P04018	貨郎	300025250	peddlers	Н	H.HG
31	P04019	行旅	300025879	travelers (G)	н	H.HG
32	P04022	獵人	300025619	hunters (people) (G)	Н	H.HG
33	P04023	樂隊	300205065	bands (ensembles) (G)	Н	H.HG

Results of Mapping Types Between NPM and AAT

Code	Code Mapping Type Quantity (percentage		Example
BM	Hierarchical (narrower to broader)	1327 (63.19%)	剔犀 (ti xi) BM lacquering (A12067)
=EQ	Exact simple equivalence	638 (30.38%)	仰韶文化 (yang shao wen hua) =EQ Yangshao (A02002)
EQ	Cumulative compound equivalence	52 (2.48%)	塔(ta) <mark>EQ</mark> pagodas towers (A11331)
~EQ	Inexact simple equivalence	37(1.76%)	后妃 (hou fei) ~ <mark>EQ</mark> empresses (P04003)
EQ+	Intersecting compound equivalence	33(1.57%)	春景 (chun jing) <mark>EQ</mark> springs + scenes (P03001)
NM	Hierarchical (broader to narrower)	9 (0.43%)	玉石器 (yu shi qi) NM jades (objects) (A04004)
RM	Associative mapping	4 (0.19%)	臺閣 (tai ge) RM pavilions (P14002)
總計		2100(100%)	

Interoperability of Conceptual Structures for NPM in AAT



		Structures				
		Specific category in AAT	No specific category in AAT			
	Completely covered in AAT	Pattern 1 A02 考古學文化 (13 values) Archaeological Culture A17 時代 (66 values) Chinese Periods				
Vocabularies	Incomplete covered in AAT	Pattern 2 A03 窯系 (19 values) Chinese Ceramic Style	Pattern 3 A08 釉色 (69 values) Chinese Ceramic Glazes P02 中國繪畫技法 (34 values) Techniques of Chinese Painting			
	None covered in AAT		Pattern 4 P17傳說動物 (7 values) fabulous Animals P18中國節令 (8 values) Chinese Festivals			

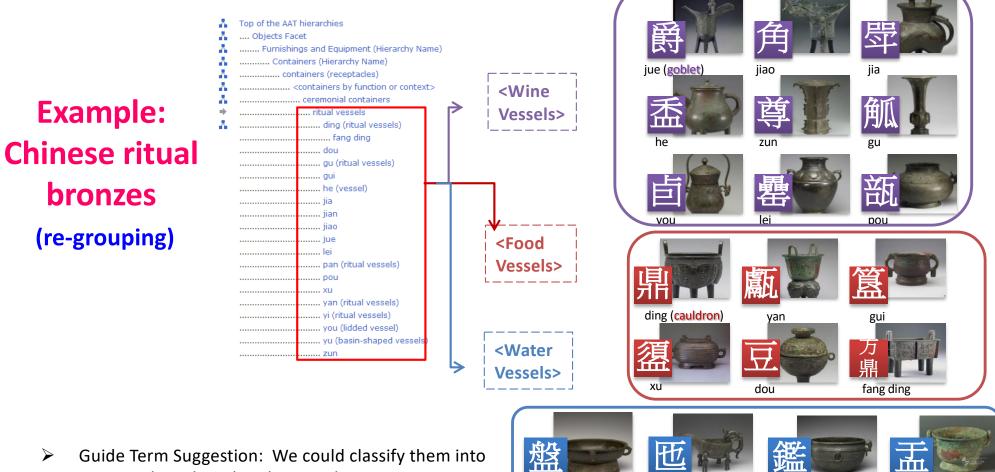
Conceptual structures for the Chinese art concepts covered in AAT - Four Patterns

- 1. The values listed in a unit of the NPM are completely covered by the corresponding AAT category.
- 2. The values listed in a unit in NPM are not completely covered by that AAT category.
- 3. The values in a unit in NPM are across in various places in AAT.
- 4. No specific category and equivalent concept in AAT for the Chinese art concepts.

The **Concept** Issues

- Chinese-Specific Culture
 - Create a Whole Set of Terms (Chinese Scripts)
 - Re-Group Terms (Chinese Bronze Vessels)
 - Term Disambiguation (Chinese Festivals)
 - Lantern Festival
- Western-Specific Culture
 - The Loanword Principle
 - Term Disambiguation (Translated Terms)
 - Principle 1. Add Qualifiers to Chinese Terms (The same terms exist when translated to Chinese). Chinese qualifiers are added according to the priority rules below:
 - Upper contexts
 - Combination of multiple parent nodes
 - Extracted from scope note
 - Principle 2. Remove Qualifiers from Chinese Terms (Different terms exist when translated to Chinese.)
 - Chinese qualifiers are removed according to the elimination of homograph

AAT has these 19 ritual vessels...



pan (basin)

yi

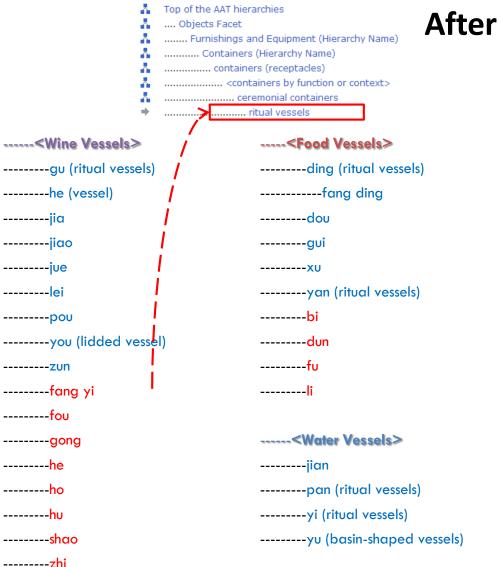
jian

yu

3 groups based on their historical usages.

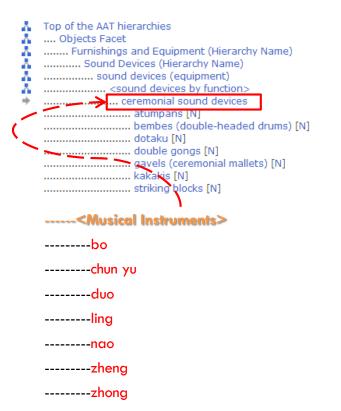


Need to add 19 NEW Chinese Ritual Bronzes!



After completion... AAT will have 38 terms with classification

Possible Hierarchical Position for Chinese Bronze Vessels- Musical Instruments



What Are Structural Issues in an AAT Context?

- The way we classify, organize and arrange terms
- The AAT can be used to guide users in enhancing their knowledge of specific topics and concepts, and can even affect their views on and perception of these concepts.
- In terms of Chinese and English mapping, the study has noticed some definite patterns in the structural issues which arise.

Facet and hierarchical structures of AAT

*	Top of the AAT hierarchies
Δ.	Associated Concepts Facet
	Associated Concepts (hierarchy name)
	Physical Attributes Facet
Λ.	Attributes and Properties (hierarchy name)
	Conditions and Effects (hierarchy name)
	Design Elements (hierarchy name)
	Color (hierarchy name)
	Styles and Periods Facet
	Styles and Periods (hierarchy name)
	Agents Facet
	People (hierarchy name)
١.	Organizations (hierarchy name)
	Living Organisms (hierarchy name)
	agents (general) [N]
4	Activities Facet
5	Disciplines (hierarchy name)
	Functions (hierarchy name)
	Events (hierarchy name)
	Physical and Mental Activities (hierarchy name)
Ε.	Processes and Techniques (hierarchy name)
	activities (general context)
	Materials Facet
4	Materials (hierarchy name)
	Objects Facet
	Built Environment (hierarchy name)
	Components (hierarchy name)
	Furnishings and Equipment (hierarchy name)
	Object Genres (hierarchy name)
	Object Groupings and Systems (hierarchy name)
	Visual and Verbal Communication (hierarchy name)

The Structural Issue

There is a conflicting structural arrangement between Western and Chinese cultures..... *To begin with, the case of ceramic glazes*

Western v.s. Chinese

In the context of Western cultures, people tend to classify *ceramic glazes*

• by composition

alkaline glaze, blue-and-white, Bristol glaze, celadon, lead glaze

- **by form** clear glaze
- by technique blister glaze, crystalline glaze, matte glaze, overglaze

Six Conjoined Vases with Tea Dust Glaze Ch'ien-lung Reign (1736-1795) Qing (1644-1911) 清 乾隆 茶葉末六聯瓶



From a Chinese perspective, curators and research studies tend to classify *ceramic glazes*

• by color

under different colors with varying compositions, forms and techniques



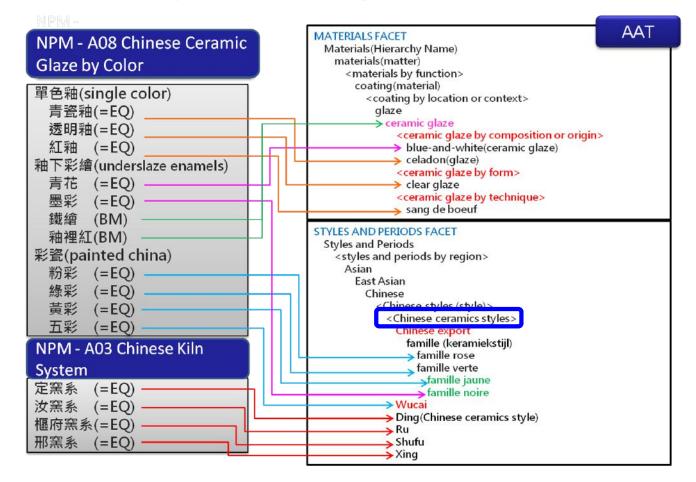
Monk's Cap Ewer with Ruby Red Glaze, Hsüan-te Reign (1426-1435) Ming (1368-1644) 明宣德 寶石紅僧帽壺

The Problem Includes

- Some terms are shared by Western & Chinese cultures. Where should these terms be located within the hierarchical structure of the AAT?
- Some terms are arranged in different FACETS by Western & Chinese cultures

Ceramic Glaze by

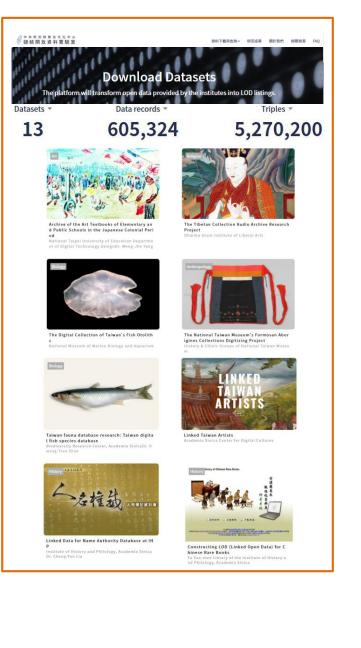
color vs. composition/origin vs. form vs. technique



AAT Applications: Linked Open Data (LOD) Supporting Digital Humanities

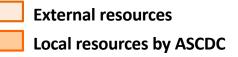
AAT as LOD(Linked Open Data) and its application in ASCDC

- Online portals for LOD Datasets, launched since 2018
- ASCDC LOD Datasets (http://data.ascdc.tw/)
- ASCDC LODLAB (<u>http://lodlab.ascdc.tw/</u>)
- Current results
- Datasets: 13
- Data records: 605,324
- Triples: 5,270,200
- Topic types of ASCDC's LOD datasets
 - Anthropology (Indigenous people in Taiwan)
 - Arts (Art Textbooks of Elementary and Public Schools in the Japanese Colonial Period, Archives of Chen Cheng-po's Paintings and Document, Linked Taiwan Artists)
 - Audio-video files (Educational and Cultural Programs from Chinese Television System)
 - Biology (Taiwan fish species database, Taiwan's Fish Otholits, Aquatic Animals in Taiwan)
 - History (Chinese Rare Books, Database for Names and Biographies, Database of Qing Official Titles, Wooden Slip Character Dictionary)
 - Religion (Tibetan Collection Audio Archive)



Data Enrichment

Overview on the reused vocabularies in ASCDC LOD datasets



	Biology	Anthropology	Religion	Arts	History	Audio-Video
Darwin Core						
DBpedia						
DC/ DCterms			I			
EDM	I					
GeoNames						
AAT						
TGN						
ULAN						
LOC						
Linked Open Copac and Archives Hub(LOCAH)						
Temporal Gazetteer (TGAZ, Harvard)						
VIAF			I			
ASCDC						
Linked Taiwan Artists (LTA)						
Database for Names and Biographies (DNB)						

Data Enrichment

Cross-dataset queries as strategy for data enrichment

HOWEVER,



• To link external resources is only one step for data enrichment, but to further reuse the data content of the linked external resources can further enlarge knowledge.

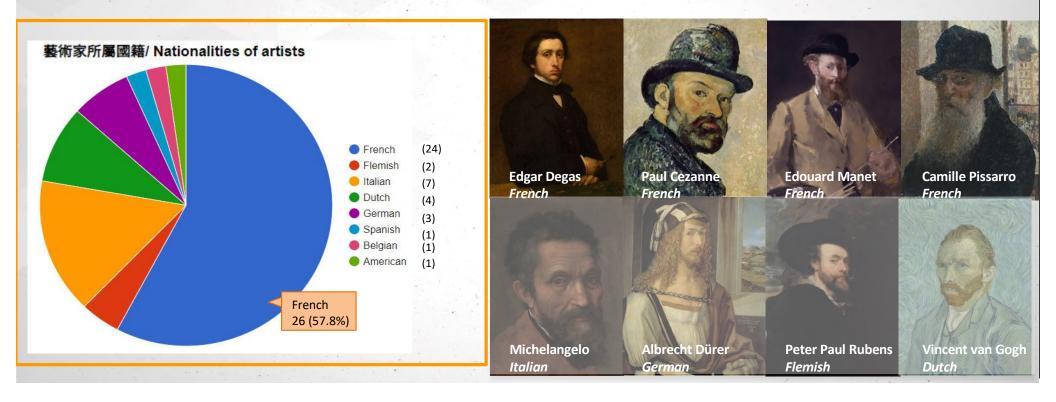
AAT as LOD Resource for Cross-dataset Queries



Getty Vocabularies: LOD

How can Linked Open Data enhance research? Data Enrichment Example of Cross-dataset Query for Knowledge Expansion

Dataset: The Archival fonds of Chen Cheng-po's paintings and documents Query: Which Western artists appear in Chen's scrapbooks? What were their nationalities and major activities?



Thanks for your attentions