CrossMark

Alignment of conceptual structures in controlled vocabularies in the domain of Chinese art: a discussion of issues and patterns

Shu-jiun Chen[1] · **Marcia Lei Zeng**[2] · **Hsueh-hua Chen**[3]

Abstract In any effort that involves multilingual KOS, the language and cultural variants bring a great degree of complexity, especially when perspectives of different cultures need to be integrated. This paper attempts to explore, within the domain of Chinese art, the interoperability of multilingual/multi-cultural vocabularies by analyzing the issues of aligning the conceptual structures of Chinese and English terminology in the controlled vocabularies observed in our Chinese AAT-Taiwan Project (2010-present). The paper addresses the issues related to the conceptual structures of the concepts for Chinese art in the National Palace Museum (NPM) Vocabulary and Art & Architecture Thesaurus (AAT). The paper presents four patterns found in the effort of achieving semantic interoperability and shares a set of models that have been used to handle various situations. The findings help the understanding of the semantic interoperability of multilingual KOS, especially when dealing with cultural-related concepts that cannot be exactly aligned in vocabularies due to discrepancies in the conceptual structures.

✉ Shu-jiun Chen
sophy@sinica.edu.tw

Marcia Lei Zeng
mzeng@kent.edu

Hsueh-hua Chen
sherry@ntu.edu.tw

[1] Institute of Taiwan History, Academia Sinica, Taipei, Taiwan

[2] School of Library and Information Science, Kent State University, Kent, OH, USA

[3] Department of Library and Information Science, National Taiwan University, Taipei, Taiwan

1 Introduction

Semantic interoperability of knowledge organization systems (KOS) is discussed in literature related to information services at all levels due to its direct impact on the quality of searching and browsing performed by users. Semantic interoperability is the capacity for different agents, services, and applications to communicate data, information, and knowledge while ensuring accuracy and preserving meaning [21]. In any effort that involves multilingual KOS, the language and cultural variants bring a great degree of complexity, especially when perspectives of different cultures need to be integrated. This paper attempts to explore, within the domain of Chinese art, the interoperability of multilingual/multicultural vocabularies by analyzing the issues of aligning the conceptual structures of Chinese and English terminology in the controlled vocabularies observed in our Chinese AAT-Taiwan Project [8]. Here, "conceptual structure" means the semantic context of a concept such as the hierarchical structure or category, in which a concept is situated within a thesaurus. Although in the project we also completed the major mapping process for the Chinese and English terms representing concepts [5,6], this paper will only address the issues related to the conceptual structures of these concepts for Chinese art in the National Palace Museum (NPM) Vocabulary and Art & Architecture Thesaurus (AAT). The ultimate goal of the paper is to explore the extent of semantic interoperability of Chinese art-related concepts in current Western art-based thesauri, as reflected particularly through the discrepancies in the conceptual structures of concepts in the NPM Vocabulary and AAT.

Unit			Value Space
Elements	**Sub-elements**		**代碼表 (Controlled values - first level)**
考古學文化 (Archaeological Culture)			河姆渡文化｜仰韶文化｜半坡類型｜廟底溝類型｜馬家窯文化｜馬家窯類型｜半山類型｜馬廠類型｜大汶口文化｜紅山文化｜良渚文化｜龍山文化｜齊家文化｜其他文化類型
窯系 (Kiln System)			越窯系｜邢窯系｜汝窯系｜哥窯系｜樞府窯系｜宋官窯系｜明代官窯系｜清代官窯系｜定窯系｜磁州窯系｜耀州窯系｜景德鎮窯系｜龍泉窯系｜鈞窯系｜遼窯系｜建窯系｜吉州窯系｜霍窯系｜德化窯系｜其他窯系
尺寸與重量 (Dimension)	檢索尺寸 (Size)	位置 (Location)	全器｜器身｜外壁｜內壁｜面｜背｜兩側｜兩端｜首｜尾｜蓋｜蓋（外）｜蓋（內）｜口｜口（外）｜口（內）｜頸｜肩｜把｜執｜耳｜提梁｜流｜腹｜腹（外）｜腹（內）｜底｜底（內）｜底（外）｜脛｜足｜足（內）｜足（外）｜座
		類別(Type)	高｜長｜寬｜厚｜徑｜深｜重量｜寬厚指數
		尺寸單位 (Value)	公分｜公克
類別 (Genre)			銅器、金屬器｜陶瓷器｜玉石器｜木竹漆器｜牙骨角蚌皮器｜琺瑯器與玻璃器｜其他
功能 (Function)			盛裝器｜食器｜飲器｜水器｜雜器｜服飾、佩飾｜耳飾｜手飾｜成扇｜家具｜照明器｜香具｜陳設｜花器｜工具｜建材｜車馬器｜文具、文玩｜印璽｜書｜畫｜樂器｜法器｜明器｜禮器｜量器｜貨幣｜兵器｜功能不明｜其他功能
形式描述 (Physical Description)	描述類別 (Description Type)		質材(material)｜胎質(ware)｜釉色(glaze)｜硬度(hardness)｜色澤(color)｜成份(component)｜全器形制 (main shape)｜局部形制｜製作痕跡｜使用痕跡與遺存

Fig. 1 Extracted content from the NPM Metadata Requirement Specifications, indicating the elements, sub-elements, and controlled values. English headings added by authors

2 Background of the vocabularies

As one of the top five museums in the world, the National Palace Museum (NPM) in Taipei holds a large collection of the most precious items of Chinese art. In the registration and description of its over 80,000 items, NPM implements a set of metadata requirement specifications. The controlled values for each metadata element in the specifications are provided by the NPM; together, there are over 2000 Chinese terms representing concepts [1,2]. Although these sets of controlled lists have never been published as a thesaurus like AAT, they have been designed, maintained, and used as a unified controlled vocabulary for the museum's collection. Hence, this research considers the whole set of the controlled lists specified by the NPM Metadata Requirement Specifications as the NPM Vocabulary when conducting the alignment task.

Figure 1 presents an extracted section from the NPM Vocabulary, indicating the elements, sub-elements, and controlled values of the first level (Fig. 1, English translations in all figures are added by authors). For the element "Function" of artifacts (see the rectangle-marked element in Fig. 1), the pre-defined values include *drinking vessel, costume,* and *musical instrument*, among others.

Another example is the element "Technique," regarding calligraphy and painting, and the pre-defined values include *baimiao* (白描, a technique executed in fine lines and delicate colors), *gongbi* (工筆, a technique using meticulous brush work with attention to detail, form, and standard), and *xieyi* (寫意, a method using free sketch or freehand brushwork). These will be used as examples in later sections of this article.

The values assigned for different elements may extend as deep as four levels in a hierarchy. The grouping and hierarchical structure setting are based on the longtime experiences and best practices of Chinese art research. Take the example of "Physical Description" element; the value circled in Fig. 1, *main shape*, has three more levels of hierarchy ([*main shape*]→ *shape of an animal → shape of a four-footed animal → shape of a cow/ox/bull*) (Fig. 2).

AAT is a multilingual thesaurus developed and maintained by the Getty Research Institute in the USA with over 250,000 terms on art and architectural history, styles, and techniques.

| Element | Examples of Hierarchical Levels in the NPM Vocabulary | | | |
	1st Level	2nd Level	3rd Level	4th Level
	xxx	. xxx	. . xxx	. . . xxx
色澤 Color	水墨 monochrome ink			
主題 Subject	山水 landscape	春景 spring scene		
質材 Material	植物 plant	果實 fruit	葫蘆 gourd	
形式描述 Physical Description	全器形制 main shape	動物 shape of an animal	走獸式 shape of a four-footed animal	牛形 shape of a cow/ ox/ bull

Fig. 2 Examples of hierarchical levels of terms in the NPM Vocabulary

Currently, it has over 35,000 concepts, each represented by English and other language terms (including Spanish and Dutch terms for each concept, French and Italian terms for a small number of concepts, and increasingly added Chinese and German terms). Since February 2014, the Getty has released the AAT as Linked Open Data (LOD) under an Open Data Commons Attribution License (ODC-BY 1.0) to make art and research resources as widely accessible as possible [13]. AAT began with Western content and has expanded continually to include other cultures from around the world. AAT is constructed according to the ISO standard for thesauri [16] and always considers semantic relationships (hierarchical and associative relationships) between concepts and equivalence relationships of terms representing the same concept [12]. The hierarchical relationships include genus/species, whole/part, and class/instance relationships between and/or among concepts. AAT organizes all concepts within seven facets: *Associated Concepts, Physical Attributes, Styles and Periods, Agents, Activities, Materials,* and *Objects* (Art & Architecture Thesaurus Online, [3]). A new *Brand Names* facet was established recently and has not been covered by this study. There are well-established hierarchies for all concepts in the thesaurus. For example, in the *Styles and Periods Facet*, concepts are presented in clusters according to general era (ancient, prehistoric, pre-ceramic, etc.), region (Asian, Early Western World, etc.), and generic styles, periods, and cultures (antique, contemporary, southwest, etc.). Almost all Chinese painting styles are organized under the concept "Chinese painting styles" (in *Styles and Periods Facet)*, with a total of 17 concepts (including *Bai miao*) listed (left side of Fig. 3). On the other hand, *gongbi* and *xieyi* are grouped under the guide term ⟨painting techniques by application method or circumstances⟩, which falls under the *Activities Facet* (right side of Fig. 3).

The need for aligning the NPM Vocabulary with AAT became significant in the digital age. In a large national digital archive initiative spanning from 2002 to 2012, the Taiwan e-Learning and Digital Archives Program (TELDAP) exposed to the world the digital representations and metadata of five million cultural objects, including thousands held in the National Palace Museum, through the unified platform of the Taiwan Digital Archives (http://culture.teldap.tw/culture/, http://digitalarchives.tw/). For the first time, people from the entire world could see these treasures without geographic barriers. However, effectively accessing these resources required further efforts to help users overcome the language and cultural barriers. The AAT's translation (http://aat.teldap.tw/) by the AAT-Taiwan sub-project of TELDAP has greatly enabled cross-language searching.

Table 1 provides an overview of the statistics of the mapping between the terms representing concepts in the NPM Vocabulary (as the source vocabulary) and those in AAT (as the target vocabulary). The framework for mapping types and the codes were defined by the ISO standard ISO25964-2 *Thesauri and Interoperability with other Vocabularies—Part 2: Interoperability with other Vocabularies* [18], which was built on the earlier ISO, NISO, and British standards [10]. About one-third of the involving NPM concepts were considered as equivalence mapping. In addition, the majority of mapping results fell into "narrower to broader" mapping type, or "BM", reflecting the fact that AAT has seven facets and many sub-facets, hierarchies, and clusters, so that a special Chinese concept could always find a broader placeholder or "home".

Nevertheless, the conceptual structures of AAT and the NPM Vocabulary revealed both linguistic and cultural preferences and differences. The NPM Vocabulary was initially developed for registering and managing physical objects; the-

Fig. 3 Extracted sections from AAT depicting where *Bai miao* and *gongbi* are presented

sauri like AAT were designed to control vocabularies and to be used in indexing and accessing all kinds of resources. Such fundamental differences also added challenges of aligning the contents of these vocabularies.

3 Research questions

When aligning the NPM Vocabulary with AAT—as in any other multilingual vocabulary mapping process—we first dealt with various types of equivalence relationships between terms representing concepts. These included exact equivalence and different types of partial equivalence. In addition, we also encountered issues in aligning the terminology's conceptual structures presented in the vocabularies. As an example, the concept *baimiao* is listed under "Technique" in the NPM Vocabulary; its English equivalent, however, is categorized as a style under the *Styles and Periods Facet* in AAT rather than under "painting techniques" in *Activities Facet* (Fig. 3). This brought us a question in the alignment: Should one of the conceptual structures be prioritized over the other? If we chose AAT as the principle structure, the conceptual structure of "Technique" in NPM would become dis-aligned, which could result in a loss of integrity in presenting the knowledge of Chinese art. Meanwhile, if we could not keep up with the conceptual structure defined by AAT, then the whole point of aligning with AAT would be minimized. Thus, we asked, Should we keep both conceptual structures? If so,

how could we process and achieve the alignment? In fact, this type of variance became a recurring problem. Thinking of many specialized terms in the domain of Chinese culture and art, we would ask similar questions again and again. Eventually, we asked, could each of the Chinese art concepts be located in an equivalent or similar conceptual structure in AAT? Could all of these concepts be mapped reasonably to AAT? Furthermore, could those Chinese and English terms that meet the "exact equivalence" mapping criteria be different in the conceptual structures? And finally, could there be different levels of discrepancies in their contexts?

This study attempts to answer such questions while addressing general research concerns related to the structural discrepancies in the alignment of terminology in multilingual and multi-cultural contexts.

4 Methodology

This study has followed the best practices of treating one vocabulary as the "source language" (SL) and another as the "target language" (TL) in the mapping process. The abbreviations "SL" and "TL" have been used in professional translations, therefore we adopted them for easier communicating among the participates who came from various disciplines. The SL used in this research is the NPM Vocabulary, and the TL is AAT. The study took about six

Table 1 Mapping types between NPM Vocabulary and AAT

Code*	Mapping Type	Num-ber	Percen-tage	Example		
				NPM term *pron.* [ID]		AAT term [ID]
BM	Hierarchical (narrower to broader)	1327	63.19%	剔犀 *ti xi* [A12067]	BM	lacquering [300053796]
=EQ	Exact simple equivalence	638	30.38%	仰韶文化 *yang shao wen hua* [A02002]	=EQ	Yangshao [300173481]
EQ |	Cumulative compound equivalence	52	2.48%	塔 *ta* [A11331]	EQ |	pagodas (buildings) [300004829] | towers (single built works) [300004847]
~EQ	Inexact simple equivalence	37	1.76%	后妃 *hou fei* [P04003]	~EQ	empresses [300150492]
EQ+	Intersecting compound equivalence	33	1.57%	春景 *chun jing* [P03001]	EQ+	spring (season) [300133097] + scenes (depictions) [300264666]
NM	Hierarchical (broader to narrower)	9	0.43%	玉石器 *yu shi qi* [A04004]	NM	jades (objects) [300184767]
RM	Associative mapping	4	0.19%	臺閣 *tai ge* [P14002]	RM	pavilions (garden structure) [300006819]
Total		2100	100%			

	*Code meanings according to ISO25964 -2 (2013):
BM	Broader mapping; the term that follows the tag represents a concept having a wider meaning.
EQ	Equivalence; the term that follows the tag is the preferred term in a target vocabulary that is closest in meaning to the preferred term preceding the tag, from the source vocabulary.
=	An equals sign is used in conjunction with a mapping abbreviation to indicate that the mapping is exact. Specifically, "=EQ" signifies exact equivalence.
|	A vertical bar or "pipe line" is shown between two or more preferred terms from a target vocabulary, the scopes of which, in sum, best cover the scope of a broader preferred term in the source vocabulary.
~	A tilde sign is used in conjunction with a mapping abbreviation to indicate that the mapping is inexact. Specifically, "~EQ" signifies inexact equivalence.
+	A plus sign is shown between two or more preferred terms from a target vocabulary that are used in conjunction to represent a compound concept in the source vocabulary.
NM	Narrower mapping; the term that follows the tag refers to a concept with a more specific meaning.
RM	Related mapping; the term that follows the tag labels an associated concept, but is not a synonym, a quasi-synonym, a broader term or a narrower term.

months. Besides the authors, the study included five Chinese art domain experts and the AAT editors of the Getty Vocabulary Program (GVP), engaging them in discussions when facing uncertainty in the concept alignment (for example, when a concept could potentially be mapped to more than one concept in different hierarchies). The study encouraged different opinions among researchers about the alignment results in order to ensure the reliability and validity of those results. In addition, the study adopted three theories—literary warrant [4,15], qualia structure [20] and semantic relations [9,12,14,17]—as bases for cross-examining the study's results. Further evaluation, adjustment, and recommendations have been brought by the AAT-Taiwan team to the AAT's International Terminology Working Group (ITWG) annual meetings since 2009. Ongoing correspondences with the Getty Research Institute have also been maintained. The results can be seen in both the alignments and the enrichment of the vocabularies.

The methods and procedures of this project consist of the following:

First, the metadata elements that are unique to Chinese culture, and the values associated with them, were identified in the NPM Vocabulary. The study considered the metadata elements to which the controlled values belonged as the basic categories. This resulted in categories such as "P02 Technique," "A06 Material," and "A02 Archaeological Culture." Consequently, the study examined the controlled values listed under each NPM metadata element and identified those that have strong Chinese artistic and cultural characteristics. Each case was unique, but certain types of situations were common in the study, as presented in the following examples:

- A category may be Chinese culture-oriented. For example, NPM Vocabulary's category "A02 Archaeological Culture" refers to Chinese archaeological culture, providing thirteen values [e.g., *Yangshao Culture* (A02002)] that are all unique to Chinese culture.
- In contrast, a category may not be culture-oriented at all. Category "A06 Material" refers to the material of artifacts, under which there are 110 controlled values, most of which are not Chinese culture-dependent [e.g., *shell* (A06002), *bamboo* (A06030), and *emerald* (A06099)].
- A category may encompass both unique Chinese art concepts and culturally shared concepts. For instance, category "P02 Technique" refers to calligraphy and painting techniques, under which the thirty-five controlled values are mostly unique to Chinese art [e.g., *baimiao* (P02001), *gongbi* (P02002), and *xieyi* (P02003)], while some others are mutually shared by Chinese and Western cultures [e.g., *oil painting* (P02008)].

Second, the study conducted an analysis of the alignment of NPM Vocabulary and AAT. The focus was on the following situations and questions:

(a) For areas that are especially unique to the Chinese culture, the question arose as to whether AAT provided a specific category that matches the one in the NPM Vocabulary;

(b) After considering the findings from (a), the question arose as to whether the matching category in AAT covered the values (terms representing concepts) listed in a NPM category completely; and

(c) For those NPM categories that cannot be exactly aligned, the question arose as to how we could make both English and Chinese users understand their meanings correctly.

For instance, "A02 Archaeological Culture" in the NPM Vocabulary can be mapped to the corresponding category under the concept *Chinese Neolithic periods* in AAT, with the terms having an exact equivalence relationship (indicated as Pattern #1 in Fig. 4). However, in many other cases, this ideal situation could not be found. Such a situation can be further characterized with three patterns, as indicated in Patterns #2 to #4 in Fig. 4. As a result, the study had to combine different modules in the processes according to different mapping conditions in order to achieve the best mapping quality.

5 Findings

The study observed four patterns of alignment between the conceptual structures of the NPM Vocabulary and AAT, as summarized in Fig. 4. The numbers shown in the parentheses indicate the quantity of involved terms in the selected categories (Fig. 4). The patterns will be explained in the following sections.

5.1 Pattern #1

Pattern #1: There is a specific category in AAT corresponding to a category in NPM Vocabulary, with the values in the category of NPM also completely covered by that AAT category (e.g., NPM category for "A02 Archaeological Culture" in Fig. 5).

Take a closer look at the instance in Fig. 5, in which *Yangshao* (ID: 300173481) appears under *Chinese Neolithic periods* (ID: 300018340) in AAT. (The Yangshao culture was a Neolithic Chinese culture and period that flourished ca. 5000 to ca. 3000 BCE.) This pattern may indicate that values represented by such thematic vocabularies are shared by the East and the West, or that certain themes specific to Chinese culture have already been accommodated in AAT.

Fig. 4 Patterns of conceptual structures for Chinese art concepts found in the NPM Vocabulary and AAT mapping

		STRUCTURE	
		Specific category in AAT	No specific category in AAT
CONCEPTS	Completely covered in AAT	**Pattern#1** A02 考古學文化 (13 values) Archaeological Culture A17 時代 (66 values) Chinese Periods	
	Incompletely covered in AAT	**Pattern#2** A03 窯系 (19 values) Kiln System	**Pattern#3** A08 釉色 (69 values) Chinese Ceramic Glaze P02 中國繪畫技法 (34 values) Chinese Painting Technique
	Not covered in AAT		**Pattern#4** P17 傳說動物 (7 values) Fabulous Animals

Fig. 5 Example of Pattern #1, showing the matching NPM category and the members of the AAT category
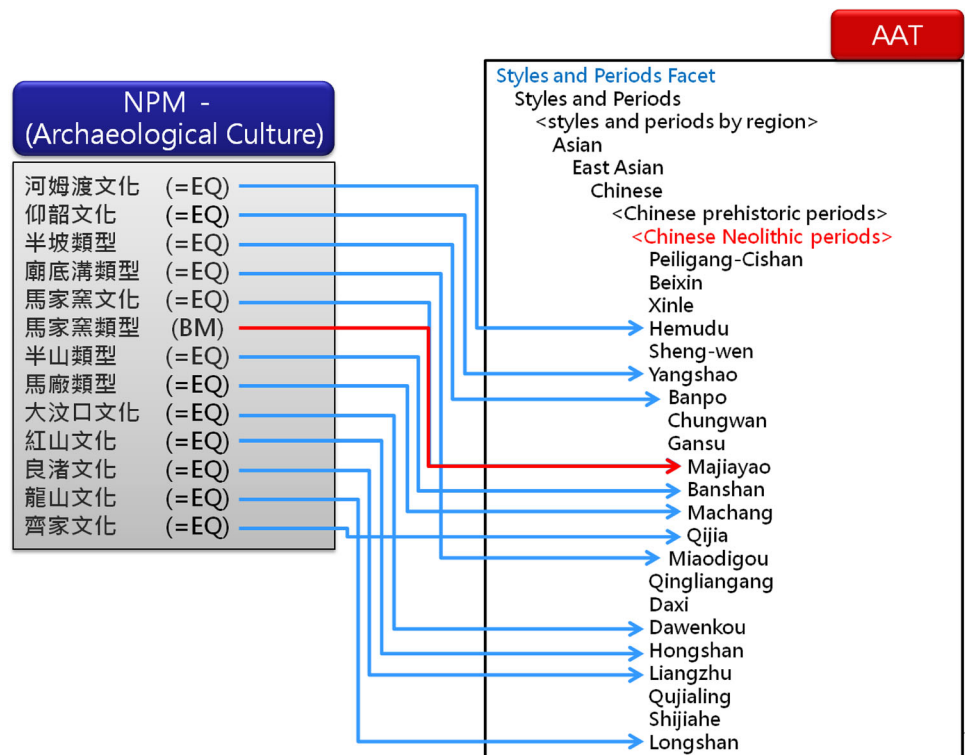
Note: " =EQ" in the left exhibit refers to the Exact simple equivalence; "=BM" refers to the Hierarchical mapping (a narrower concept mapped to a broader concept).

Another example is for the NPM category "A17 Chinese Period," for which AAT lists the *Tang (Chinese style)* under *Chinese dynastic styles and periods*.

Pattern #1 is thus the ideal pattern in which the concepts and structures of two different structured vocabulary systems are highly interoperable. However, as also observed in the study, certain correlations in terms of semantics, development, or inheritance can be found by tracing notes of different vocabularies not yet officially established in AAT. For example, relations with respect to development and inheritance can be identified in the notes of *Yangshao* (ID: 300173481, a Neolithic Chinese culture and period that flourished ca. 5000 to ca. 3000 BCE.) and *Longshan* (ID: 300018347, a Chinese Neolithic culture and period spanning ca. 2500 to ca. 2000 BCE.). AAT has not established any kind of link between the two terms to indicate their relationships. Nevertheless, we found that AAT has already created links between other similarly correlated terms. For example, the relationship between *First Dynasty (Egyptian)* (ID: 300020269) and *Second Dynasty (Egyptian)* (ID: 300020271) are clearly identified using "preceded" and "followed" relationship types: the former "preceded" the latter and the latter "followed" the former. In addition, it is also noted that NPM

often structures concepts in a flat, non-hierarchical way as compared with AAT; for instance, concepts under *Chinese Neolithic periods* of the NPM belong to the same hierarchical level while those in AAT are structured in three hierarchical levels with hierarchical relationships between them clearly indicated.

5.2 Pattern #2

Pattern #2: Although a matching category in AAT can be found, the values in a NPM category are not completely covered by that AAT category (e.g., "A03 Kiln Systems," as shown in Fig. 6).

This pattern indicates some discrepancies. For example, NPM's "A03 Kiln System" category corresponds to the AAT concept of *Chinese ceramics styles* (ID: 300018504). However, of the nineteen values under A03, several are not included in AAT. The study further conducted a comparison between the values specifically categorized under AAT *Chinese ceramics styles* and NPM "A03 Kiln System" to explore the differences between them. We found that in the system of AAT, *Chinese ceramics styles* category seems to cover a larger range of values than NPM "A03 Kiln System."
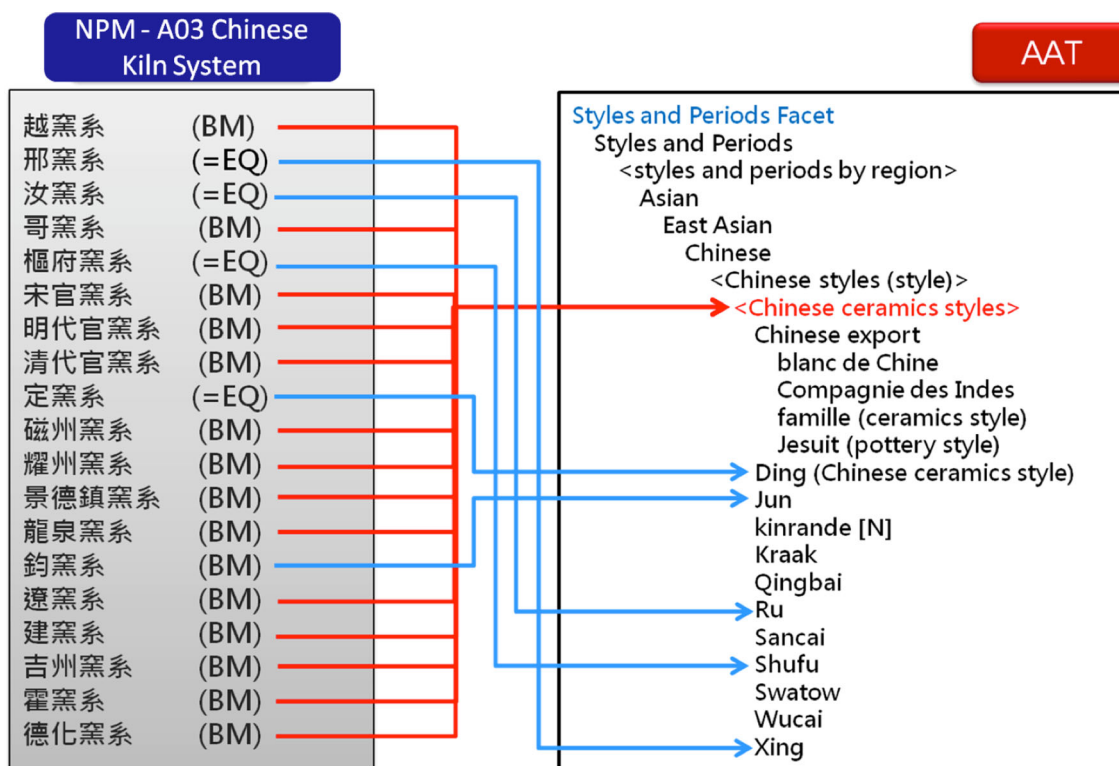
Note: " =EQ" in the left exhibit refers to the Exact simple equivalence; "=BM" refers to the Hierarchical mapping (a narrower concept mapped to a broader concept).

Fig. 6 Example of Pattern #2, showing the matching NPM category and the incomplete list of members in the category found in AAT

Fig. 7 Example of Pattern #3, showing Chinese art concepts being spread across different basic facets, hierarchies, and sub-facets in AAT
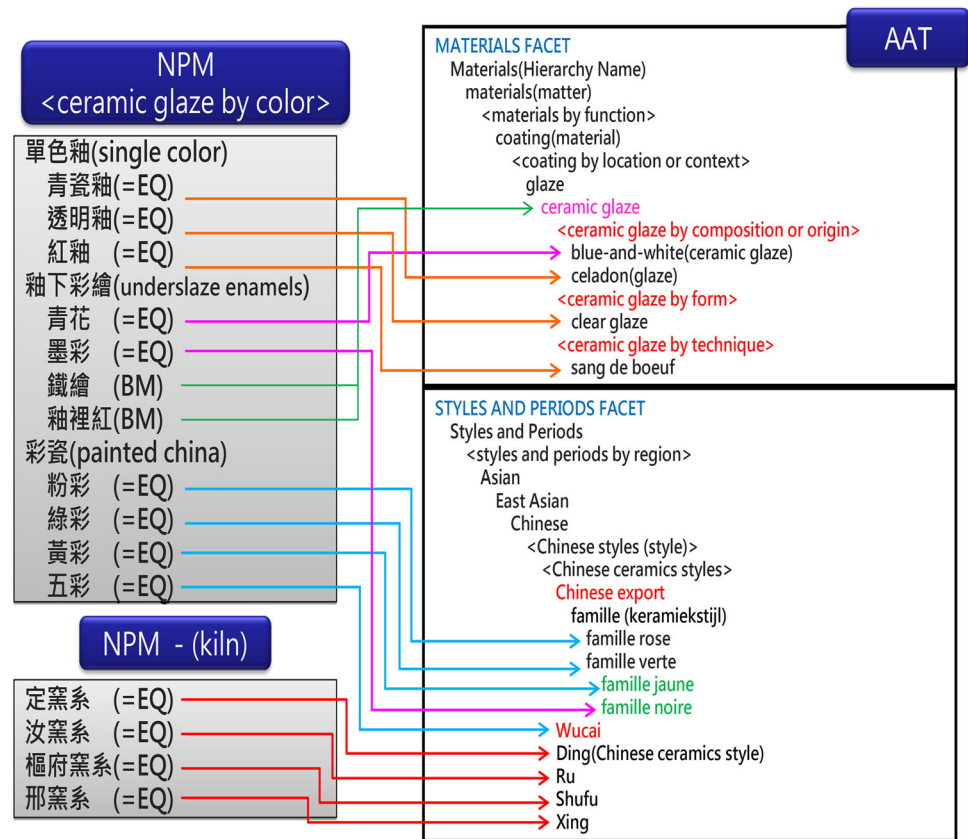
Note: " =EQ" in the left exhibit refers to the Exact simple equivalence; "=BM" refers to the Hierarchical mapping (a narrower concept mapped to a broader concept).

The differences consist mainly of terms associated with the category of *Chinese export porcelain* (ID: 300387409); AAT defines this as "Porcelain ware that is made and decorated in China to European order, as distinct from porcelain in native taste" (AAT http://vocab.getty.edu/aat/scopeNote/114840). Chinese export porcelain contributed significantly in the history of Western art and became an essential resource for the West to approach and study Chinese porcelain. However, Chinese export porcelain was not the focus of the collections in the NPM, which featured imperial collection of fine porcelains. NPM established its system of vocabularies exclusively on the basis of its collection [19].

The above example can be further explained in several situations in relation to the Chinese export porcelain:

1. With regard to time period and geographical context, the Chinese export porcelain defined in AAT was exported to Europe and later to North America during the late Ming Dynasty to Qing Dynasty. Therefore, we can deduce that AAT's coverage of *Chinese export porcelain* may not be the most complete.
2. With regard to Chinese ceramic styles, the coverage of the terms in the NPM Vocabulary has a limit as well.

NPM Vocabulary is impacted by the NPM's collection principles. The Palace Museum has implemented the "guarantee works principle" for imperial collections or specifically for the imperial use of high quality objects. Export porcelain is not in the range of collections that NPM has dedicated itself to.

To discuss the completeness issue found in Pattern #2, we can use an example from the city of Jingdezhen, known as the "Porcelain Capital" due to the quality pottery it produced for over 1700 years. Jingdezhen's porcelain was divided into three types: (1) porcelain produced in official kilns for the imperial families; (2) porcelain fired in domestic kilns for general use; and (3) export porcelain, or Kraak, that was created specifically for export. Therefore, the so-called Jingdezhen kiln system, in terms of the Central-Plains, seems to encompass the above three types. NPM's creation of the term *Jingdezhen kiln* represents its point of view and only refers to the first meaning, i.e., official kilns. As for the third one, Kraak (AAT ID: 300018509, refers to the style of porcelain made in China which formed part of the cargo of Portuguese carracks captured on the high seas by the Dutch in 1600 and taken to Holland for sale), it appears in AAT under *Chinese ceramic styles*, independent of *Chinese export*
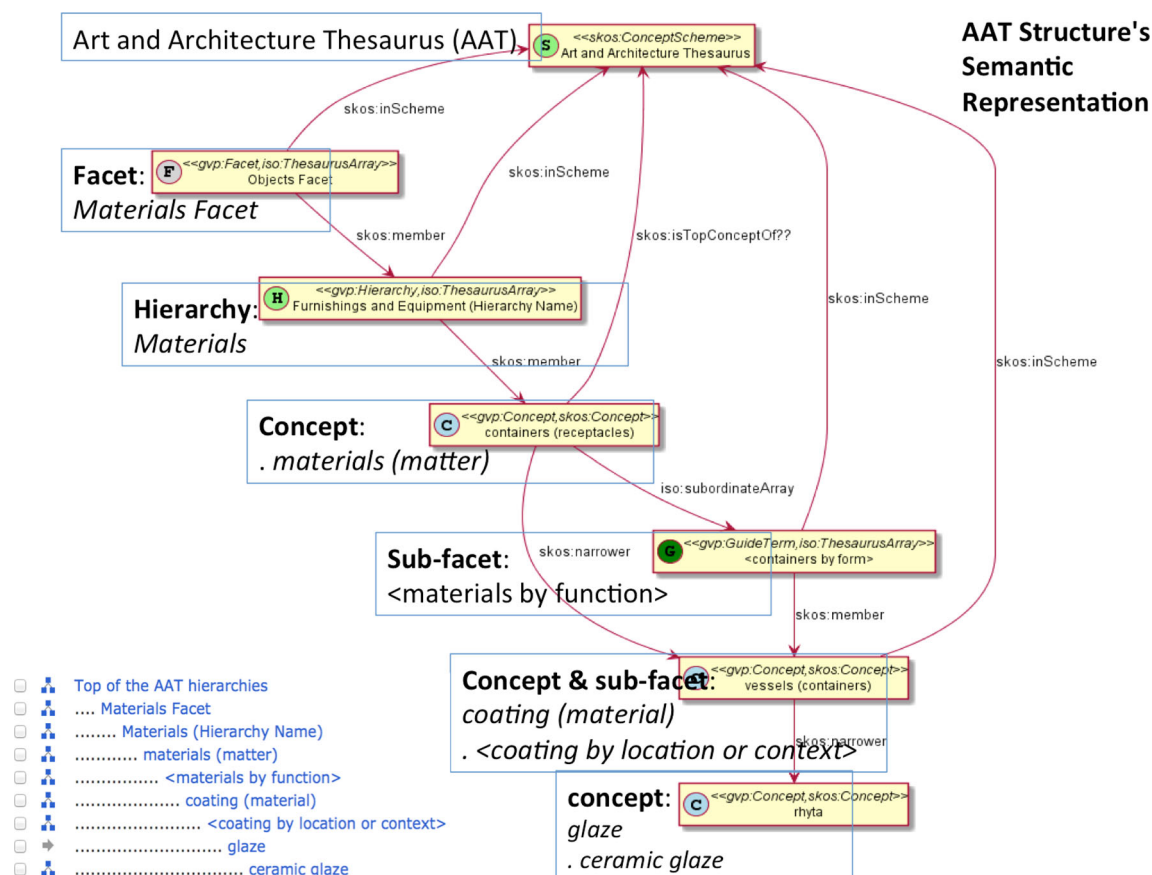
Fig. 8 AAT structure's semantic representation with examples. (Constructed based on "Getty Vocabularies: Linked Open Data Semantic Representation—Section 2.3.4 Top Concepts," 2014. Available at: http://vocab.getty.edu/doc/#The_Getty_Vocabularies_and_LOD)

porcelain. This demonstrates the complicated relationships yet to be captured in the established structure in AAT or any related vocabulary. Moreover, "Kraak" later influenced many well-known European ceramic styles such as Holland's "Delft." This term is presented in the AAT (ID:300106361), with a scope note reading: "Refers to the style of Dutch tin-glazed earthenware produced in Delft beginning in the mid-17th century and characterized by blue and white patterns inspired by imported Chinese porcelain. It also refers to similar Delft inspired wares produced in Germany and England" (http://vocab.getty.edu/aat/scopeNote/47939). Similar to what was demonstrated in the previous case, the complicated relationships that exist among *Kraak*, *Delft*, and even European ceramic styles in general may need to be established in AAT.

What these Jingdezhen kiln related concepts and terms present are the alignment challenges of term- and concept-coverage in the vocabularies, especially when multiple cultures are involved:

(a) The source language (SL) NPM Vocabulary has a complete coverage, but the target language (TL) AAT does not;

(b) The TL has a complete coverage, but the SL does not; and

(c) Neither SL or TL has a complete coverage.

The Chinese ceramic style previously mentioned belongs to situation #c. From both sides of the vocabularies involved in the alignment, we should be aware of other issues in addition to cultural points of view and different results. One thing worth noting is that the scope of the Chinese ceramic styles is much greater than the scope of the field of Chinese ceramic kilns.

5.3 Pattern #3

Pattern #3: There is no matching category in AAT, and the corresponding values in a NPM category are spread across various places in AAT (e.g., NPM's "A08 Chinese Ceramic Glaze" as shown in Fig. 7).

The NPM ceramic glaze in A08, as seen in Fig. 7, can be found in two facets and four different levels of AAT hierarchies: in the *Materials Facet* and under the hierarchy *Ceramic Glaze* (ID: 300015092) the terms are further categorized according to sub-facets ⟨Ceramic glaze by com-
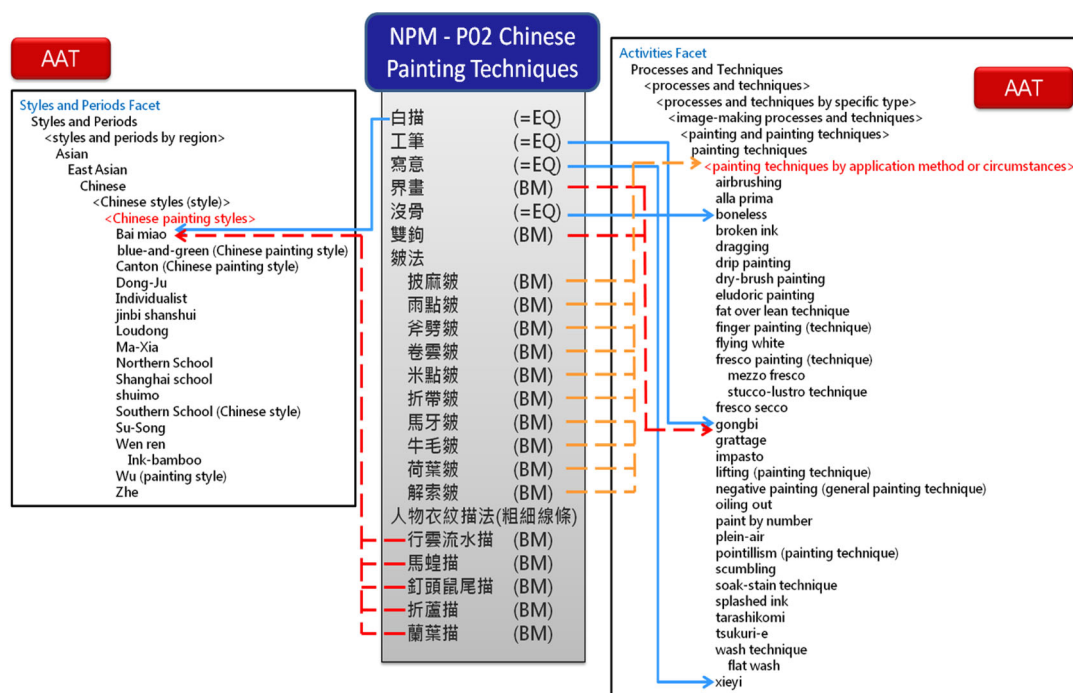
Note: " =EQ" in the center exhibit refers to the Exact simple equivalence; "=BM" refers to the Hierarchical mapping (a narrower concept mapped to a broader concept).

Fig. 9 Another example of Pattern #3, showing Chinese art concepts being spread across different basic facets, hierarchies, and sub-facets in AAT
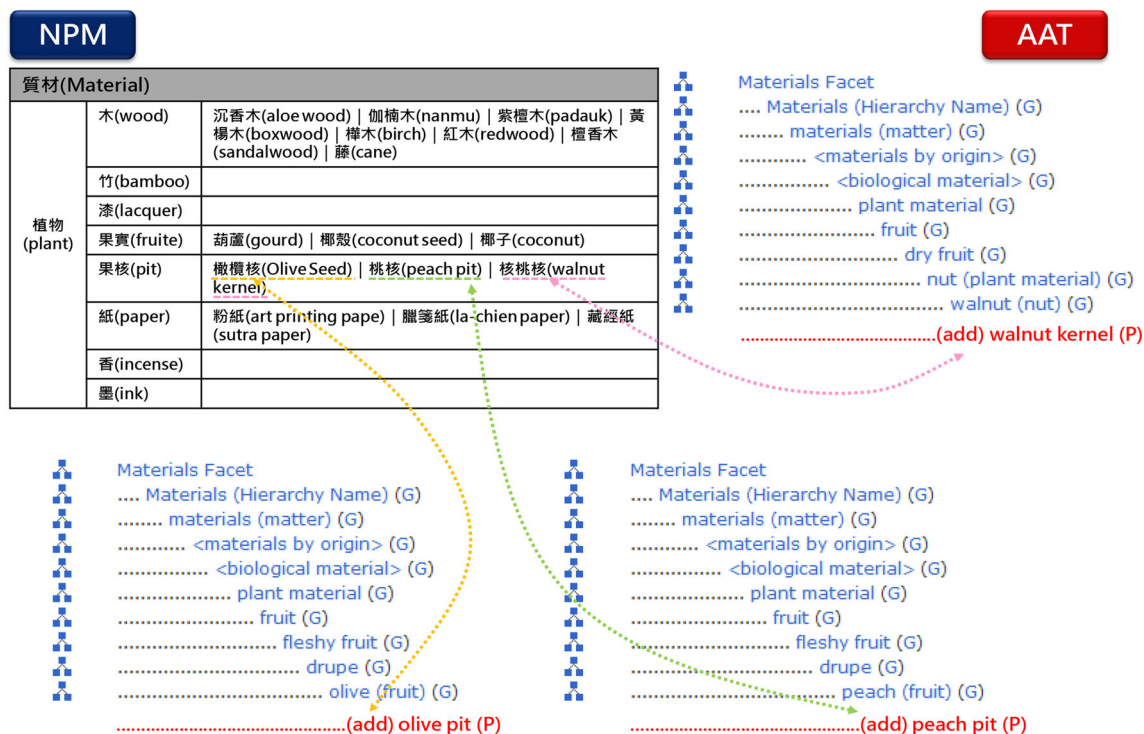
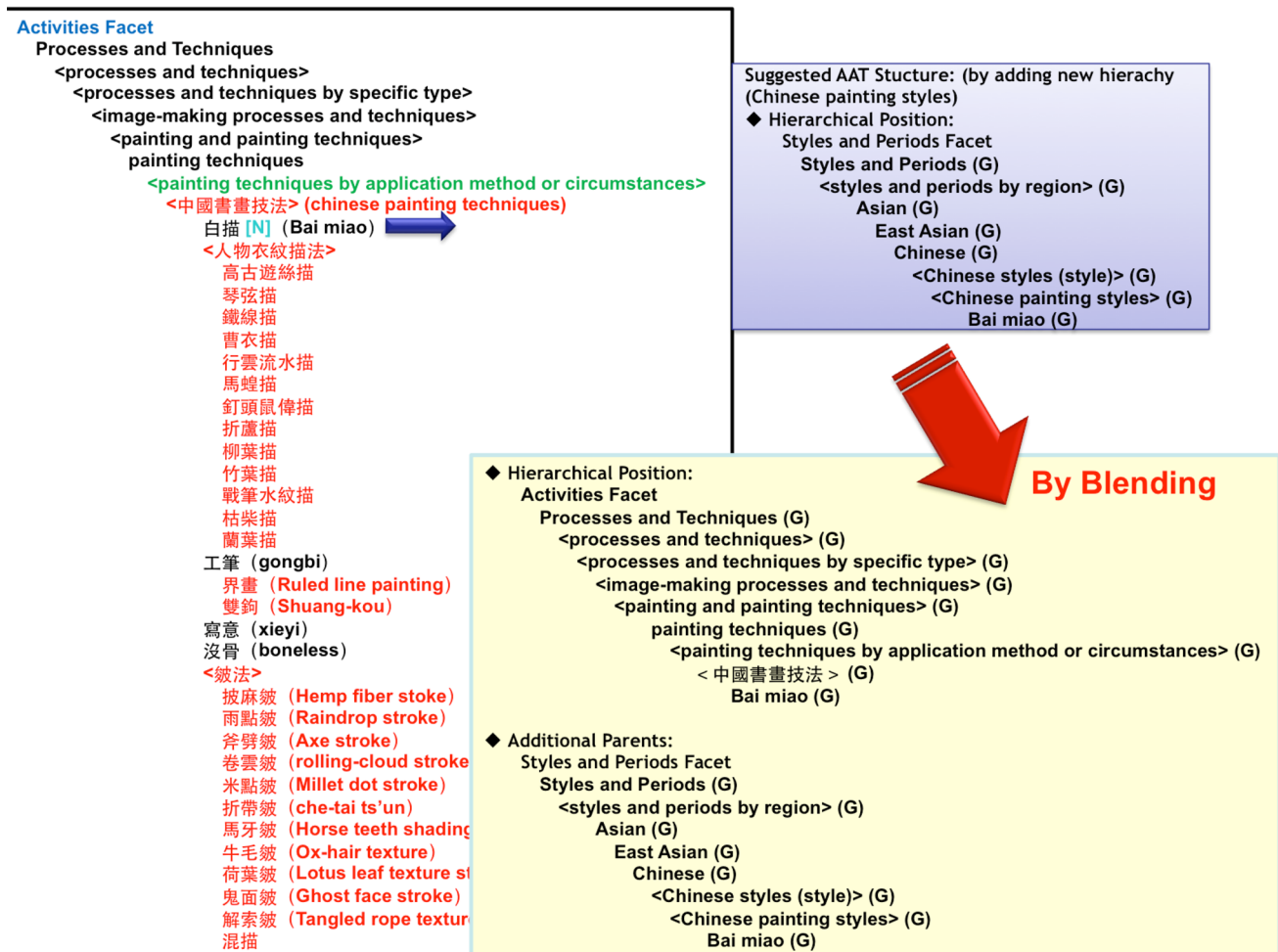**Fig. 10** Example of Model A: TL (AAT) = base; SL (NPM) = supplement

Fig. 11 Example of Model B: SL (NPM) = base; TL (AAT) = supplement

position or origin⟩, ⟨Ceramic glaze by form⟩, and ⟨Ceramic glaze by technique⟩. The rest of the terms can be found in the *Styles and Periods Facet* under *Chinese ceramics styles* (ID: 300018504). Another possibility is that, according to AAT's classification method, the glaze types in NPM can also be categorized according to color, although there is no such sub-facet so far.

This case is evidence of the structural variation of controlled vocabularies. As we have discussed in the background section, NPM Vocabulary was initially developed for registering and managing physical objects. Its structure includes a set of controlled lists, some of which are flat and some of which are enumerative classificatory which can extend as deep as four levels. AAT, on the other hand, has a comprehensive semantic structure including basic facets at the top level, hierarchical structure in all facets, and sub-facets under hierarchies. Faceted approach enables one to look at the same thing from different perspectives (Fig. 8).

The same pattern can be found in Fig. 9 regarding a group of Chinese painting techniques. The respective NPM terms

correspond to two different facets of AAT, the *Styles and Periods Facet* and *Activities Facet*, and fall under different hierarchies and sub-facets.

It is clear that for specialized terms in the domain of Chinese culture and art, there are great discrepancies between the values in the NPM Vocabulary and AAT. Such a pattern is also found for the values associated with NPM's "A12 Artifact Technique."

5.4 Pattern #4

Pattern #4: There is no specific category in AAT, and the values in a category of NPM Vocabulary find no equivalent concept in AAT.

In general, concepts that have been strongly affiliated with the Chinese culture fall into this pattern. For example:

- NPM "P17 Fabulous Animals (傳說動物) lists well-known legendary creatures found in folklore, literature, and the arts that are unique to Chinese culture.
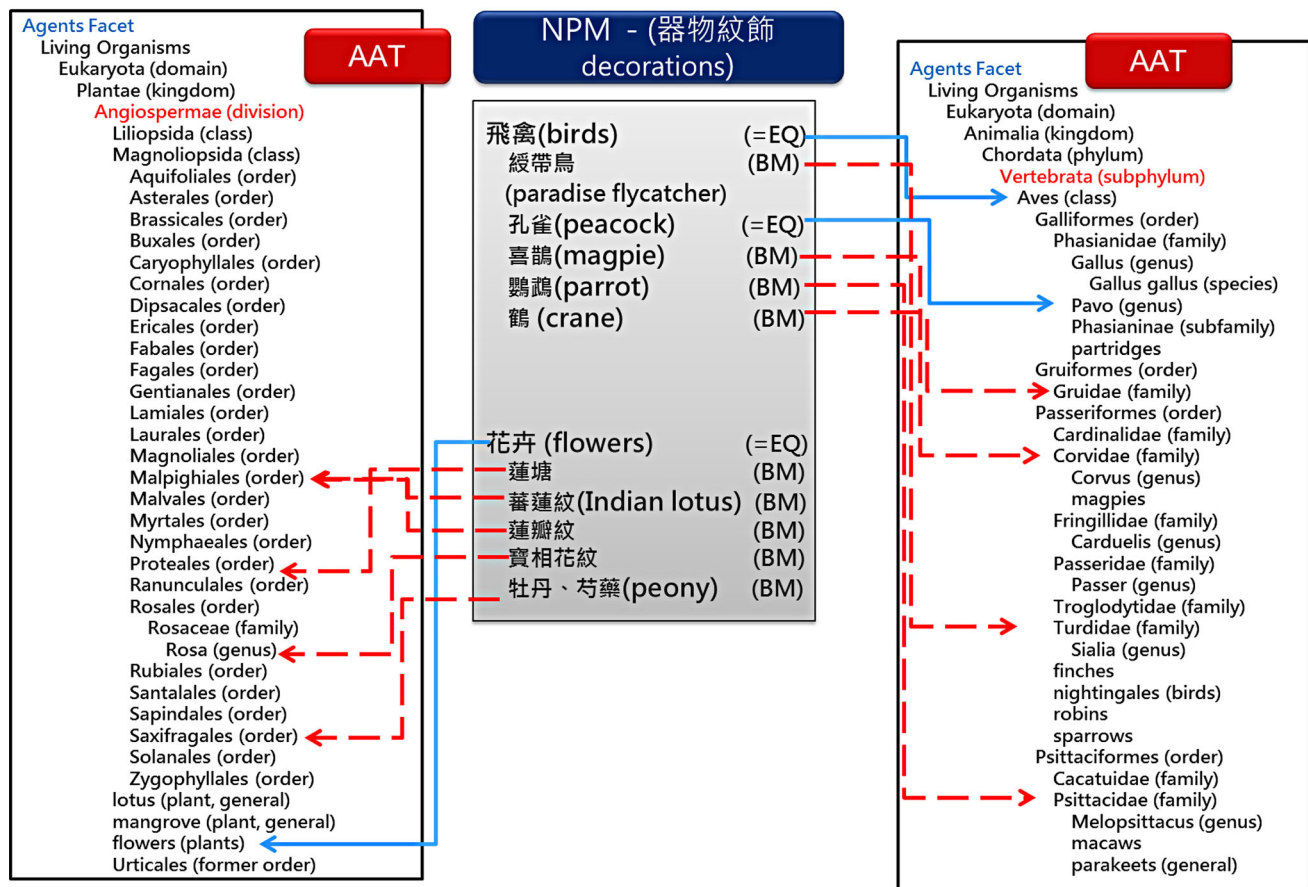
Note: " =EQ" in the center exhibit refers to the Exact simple equivalence; "=BM" refers to the Hierarchical mapping (a narrower concept mapped to a broader concept).

Fig. 12 Example of Model C: Use TL (AAT) structure

- NPM "P02-1 Texture Strokes (皴法)," a technique of using different types of brush strokes to create the textures of mountains, rocks, and trees in the Chinese painting tradition. There are 11 concepts under P02-1, including *hemp-fiber stroke, ax-cut stroke, rain dot stroke, frayed-rope stroke, lotus leaf stroke, ox furs stroke, folded-belt stroke,* and *roiling cloud stroke.*
- "P18 Chinese Festivals" lists specific Chinese festivals such as *Spring Festival, Lantern Festival, Qingming, Dragon Boat Festival, Double Seventh festival, Mid-Autumn Festival, Double Ninth Festival,* and *Spring Sacrifice* [7].
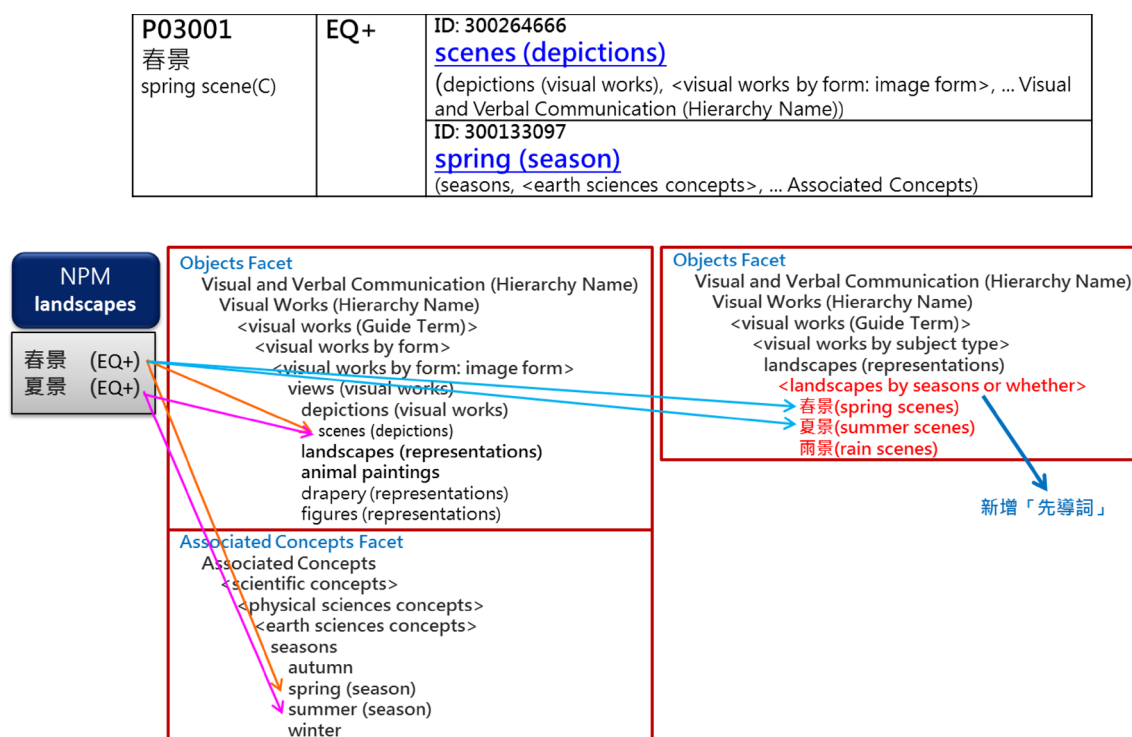
6 Discussions of the proposed models

The analysis helped the AAT-Taiwan team see the overall picture of aligning between NPM Vocabulary and AAT, while also coming up with a set of strategies for dealing with the different patterns. Four models have been established to

guide the team in proposing revisions or enhancing the overall structure. These models were not designed to respond to the four patterns, although the patterns helped develop the strategies behind them. In the alignment process, the patterns could be found to co-exist in particular categories or clusters. Thus, the proposed general models are to be implemented on a case-by-case basis:

Model A Adopting the conceptual structure of the target language (TL) as the primary structure and using that of the source language (SL) as the subsidiary structure. This is usually appropriate to the concepts that are common to both Chinese and Western culture (e.g., *plant*, as shown in Fig. 10).

Model B Adopting the conceptual structure of the source language (SL) as the primary structure and using that of the target language (TL) as the subsidiary structure. This blending mode is available in two languages, representing the unique structure and concepts of the SL (e.g., for Chinese painting techniques in Fig. 11).

| P03001
春景
spring scene(C) | EQ+ | **ID: 300264666**
scenes (depictions)
(depictions (visual works), <visual works by form: image form>, ... Visual and Verbal Communication (Hierarchy Name)) |
| | | **ID: 300133097**
spring (season)
(seasons, <earth sciences concepts>, ... Associated Concepts) |

Note: " EQ+" in the first exhibit refers to the two simple equivalences.

Fig. 13 Example of Model D: Use SL (NPM) structure

Model C Adopting only the conceptual structure of the target language (TL). The blending mode is perfectly suited for the TL (AAT) to be a more streamlined structure to show the complete vocabulary of the SL (NPM) or repetitive vocabulary. For example, the animal, plant, people, and building concepts are repeated in the NPM Vocabulary due to its non-faceted structure. AAT provided better sub-facets and hierarchies that could be adopted to overcome the limits of NPM (Fig. 12).

Model D Adopting only the conceptual structure of the source language (SL). This method is best suited for categories that carry strong cultural distinctions and works only for the source language without necessarily fitting within the target language (TL) framework. According to the analysis, this situation may also occur frequently in the case of a concept with compound terms (terms that can be split morphologically into separate components). Examples of this include an added leader term ⟨landscapes by seasons or weather⟩, which could cover *spring scenes, rain scenes, evening scenes*, etc., all of which are common themes in Chinese painting (Fig. 13)

7 Conclusion

The study reveals that AAT has already included a portion of concepts and structures that exclusively belong to Chinese arts (Pattern #1). There are still many concepts found in NPM that have not been completely covered by a corresponding category in AAT (Pattern #2), not been given a special category in AAT (Pattern #3), or not been covered in AAT (Pattern #4), even though there are major facets and hierarchies they would belong to. It is also true that the NPM Vocabulary has its preferences in terms of what concepts to include and where the concepts should be placed in the overall structure. The alignment of the conceptual structure for cultural-specific concepts needs to be handled case by case, using different models.

The analysis enabled us to understand what can be done to enhance the interoperability of both vocabularies and the virtual union catalog of the TELDEP. We started to establish semantic relationships between the concepts that have proven to be important to each other. Draft proposals for the structural changes are laid out in the translated AAT internal version.

Another important action was the collaboration with the AAT editors at the Getty Research Institute over the last several years. The AAT-Taiwan team submitted numerous proposals for adding important Chinese culture-based con-

cepts as well as their appropriate structures, keeping in mind that AAT's scope is international. The continuous collaboration, as well as the contributions of the teams for other languages from various countries, has been generating tremendous results. For example, in November 2011, a concept for *mythical or legendary beings* was created in AAT (http://vocab.getty.edu/aat/300375725). The category already included some Chinese fantasy animals (e.g., *qilin* (麒麟, ID: 3003912000), added 2014-04-07) based on our proposal. Meanwhile, some general concepts established in this category could be used in the broader matching (BM) alignment (e.g., taking *unicorn* (ID: 300379595) for *Dujiaoshou* (獨角獸)). The examples used in the figures for Patterns #2 and #3 have also been adjusted in AAT's structure. In addition, the eight terms of Chinese festivals are now added to AAT, since the study identified the lack of conceptual structure for the whole set of Chinese festivals. After the discussion at the International Terminology Working Group (ITWG) meetings and the evaluations by the AAT editors, the suggestions from the AAT-Taiwan team have been adopted by the Getty Vocabulary Program (GVP) to create a new cluster for *cultural holidays* (ID: 300400818, created 2015-04-02) under the *Activities Facet* and the *holidays* hierarchy. This demonstrates that the study results can be implemented into the real vocabulary system. The potential benefits and the collaborative results of the study include facilitating cross-lingual searching and multilingual browsing, connecting the Linked Data datasets that have used these entities, and facilitating research, discovery, and creation in the humanities.

In summary, with the efforts of the Chinese AAT-Taiwan team, the NPM and AAT vocabularies are, for the first time, put together side by side. While reviewing published literature and the international standards on thesaurus construction and mapping, the authors encountered various questions that were beyond the processes of individual concept-to-concept mapping. This paper reported issues regarding the alignment of the controlled vocabularies in the domain of Chinese art as well as the patterns found in the effort of achieving semantic interoperability. The discrepancies in the structures for many concepts that are unique to Chinese culture are significant, as reflected in the NPM Vocabulary and AAT. Such a finding calls for attention to those issues being different from *term–term translation*, *string–string conversion* and individual *concept–concept mapping* that does not consider conceptual structures. The findings in this paper are meaningful to research on the semantic interoperability of multilingual KOS, especially when dealing with cultural-related concepts that cannot be exactly aligned in vocabularies due to discrepancies in the conceptual structures.

References

1. Academia Sinica and National Palace Museum: National Palace Museum metadata requirement specification: the Painting and Calligraphy Collection. National Palace Museum, Taipei. http://metadata.teldap.tw/project/filebox/NPMshu/requirement/NPMshu_MRS_v1.0.pdf. Accessed 12 May 2015 (2001)

2. Academia Sinica and National Palace Museum: National Palace Museum Metadata Requirement Specification: The Antiquities Collection. National Palace Museum, Taipei. http://metadata.teldap.tw/project/filebox/NPMqi/requirement/npm_artifactsSpec v1.2.pdf. Accessed 12 May 2015 (2004)

3. Art & Architecture Thesaurus Online: Getty Research Institute. http://www.getty.edu/research/tools/vocabularies/aat/index.html. Accessed 12 May 2015 (2000)

4. Beghtol, C.: Semantic validity: concepts of warrant in bibliographic classification systems. Libr. Resour. Tech. Serv. **30**(2), 109–125 (1986)

5. Chen, S.J., Wu, D., Peng, P.W., Chang, Y.T.: AAT-Taiwan: toward a multilingual access to cultural objects. In: Lalmas, M., et al. (eds.) Research and Advanced Technology for Digital Libraries, pp. 389–392. Springer, Berlin, Heidelberg (2010)

6. Chen, S.J.: A study of mapping Chinese-English lexical semantics for a thesaurus in Chinese art. National Taiwan University (2012) (unpublihsed Ph.D. dissertation)

7. Chen, S.J., Kuo, C.C.: A study of knowlege organization system for digital archives in the Naional Palace Museum collections of Cheinese festivals. In: Hsiang, J., et al. (eds.) Digital Humanities and Craft: Technological Change, pp. 39–66. NTU Press, Taipei (2014)

8. Chinese AAT-Taiwan Project. http://aat.teldap.tw/. Accessed 15 Feb 2015 (2010)

9. Clarke, S.D.: Thesaural relationships. In: Bean, C.A., Green, R. (eds.) Relationships in the Organization of Knowledge, pp. 37–52. Kluwer, Dordrecht (2011)

10. Clarke, S.G.D., Zeng, M.L.: From ISO 2788 to ISO 25964: the evolution of thesaurus standards towards interoperability and data modeling. Inf Stand Q **24**(1), 20–26 (2012)

11. Getty Vocabularies Program: Getty vocabularies: linked open data. Semantic representation. Section 2.3.4 top concepts. http://vocab.getty.edu/doc/#The_Getty_Vocabularies_and_LOD. Accessed 15 Feb 2015 (2014)

12. Harpring, P.: Art and Architecture Thesaurus: Editorial Guidelines. J. Paul Getty Trust, Los Angeles. http://www.getty.edu/research/tools/vocabularies/guidelines/index.html. Accessed 12 May 2015 (2010)

13. Harpring, P.: Experiences of publishing Getty Vocabularies into LOD. http://aims.fao.org/community/interviews/experiences-publishing-getty-vocabularies-lod. Accessed 12 May 2014 (2014)

14. Hjørland, B.: Semantics and knowledge organization. Annu Rev Inf Sci Technol **41**, 367–405 (2007)

15. Hulme, E.W.: Principles of book classification. Libr. Assoc. Rec. **13**, 354–358, 389–394, 444–449 (1911)

16. ISO 5964-1985 Documentation—Guidelines for the establishment and development of multilingual thesauri. International Organization for Standardization. Revised by ISO 25964-1:2011 (1985)

17. ISO: ISO 25964-1:2011 Thesauri and Interoperability with Other Vocabularies Part 1: Thesauri for Information Retrieval. International Organization for Standardization (2011)

18. ISO: ISO 25964-2:2013 Thesauri and Interoperability with Other Vocabularies Part 2: Interoperability with Other Vocabularies. International Organization for Standardization (2013)

19. Jörg, C.J.A.: Chinese Export Porcelain. Chine de Commande from the Royal Museums of Art an History in Brussels, Exhibition Catalogue Hong Kong Museum of Art. The Urban Council, Hong Kong (1989)

20. Pustevojsky, J.: The Generative Lexicon. MIT Press, Cambridge (1995)

21. Zeng, M.L., Chan, L.M.: Semantic Interoperability. In: Bates, M.J., Maack, M.N. (eds.) Encyclopedia of Library and Information Sciences, pp. 4645–4662. Dekker Encyclopedias, Taylor and Francis Group, New York (2010)