The Getty Vocabularies' Localization and LOD Applications: Status Report from AAT-Taiwan

Dr. Sophy S.J. Chen, Lu-Yen Lu International Terminology Working Group (ITWG) meeting Los Angeles, USA. 22-24 Aug. 2016



The Getty Research Institute

花的族建築京引出







From English to Chinese

- Update on Translation Status
- · AAT-Taiwan Principle for Disambiguation





Trends in statistics

- Compared to the last statistics at ITWG 2014, numbers of users have increased (722→2,464 p. month), however pages per session have decreased (11.97→7.97 p. session). It reveals most of our current users are generally returning visitors who know the AAT structure well, and could estimate the possible place that a given concept should be located
- Phenomena can be verified that percentage of AAT-Taiwan's returning visitors are increase from 53.4% to 61.7%. The statistics show that gaining new users might be a topic to the future. The application of AAT datasets with diverse LOD projects is a possible way to gain new users.



Problems in AAT translation

• AAT translation is essentially to make a coherent and understandable lingual transformation of concepts across languages and cultures. Problems occur especially in the stage of disambiguation.

Reasons:

- Most AAT concepts originate from the Western cultures; it is hard to make a proper disambiguation for persons of other cultures, due to incoherent expressional ability between languages.
- In certain cases, meanings and texts of AAT's scope notes are quite similar, without field knowledge, it is difficult to make right disambiguation.

As a result

lower rate for disambiguation and slower contribution to AAT



- **Example:** light yellow green vs. light yellowish green
- In this case, both terms' scope notes are identical and differentiating between the two can be difficult.

AAT-Taiwan	通田	海东的夕瑶之一,长二时				
AAT-Taiwan 「通用色彩語言」中的標準色彩名稱之一,指一特定範圍的黃綠色色彩。						
AAT UCL (Universal Color Language) standard color name identifying a range of yellowish green colors.						
	identifying a range of yello	wish green colors.				
淺黃綠色 (ID: 30	00128408)	same scope notes i				
淺黃綠色 (ID: 30 light yellowish g	00128408)					
	00128408)	same scope notes in both concepts				

- **Example:** light yellow green vs. light yellowish green
- In AAT-TW, light yellow green (ID300129329) and light yellowish green (ID300128408) share the same Chinese translation as "qian huang lü se"(淺黃綠色).
- Concepts on color in AAT are structured using the classification of Universal Colors of Language-system issued by Inter-Society Color Council in the 1960s, which is fixed according to Western color theories and unfamiliar in Chinese concepts.
- According to UCL system, each color represents collections of diverse values and chroma, which makes disambiguation harder.



Solution: AAT-Taiwan Principles for Disambiguation

- Principles created according to rules of Getty's Art & Architecture Thesaurus: Editorial Guidelines.
- Main concerns for differentiation between the homographs: referring to broader context or implying significant distinguishing characteristics between terms.
- 5 major types of AAT-Taiwan Principles for Disambiguation
 - **1.** Adopting keywords in reference to Chinese scope notes
 - 2. Adopting Chinese broader terms as qualifier
 - 3. Using the Chinese preferred term as qualifier, especially for nonpreferred term
 - 4. Translating original English qualifier into Chinese
 - 5. Applying original English term as qualifier







AAT-Taiwan Principles for Disambiguation 3. Using the Chinese preferred term as qualifier Applied majorly to distinguish the non-preferred Chinese terms 編號: 300008903 詞彙類型: 概念 編號: 300343521 詞彙類型:概念 草澤 marshes 沼澤地 slews ▲..物件層面 詞彙:建築環境 (層級名) 詞彙: 沼澤地 (P.,U,繁體中文-偏好,UF,U,U) 草澤 (P,,U,繁體中文-偏好,D,U,U)(中間略) 沼澤 (草澤) (,U,繁體中文-非偏好,UF,U,U) 沼地(沼澤地) (,U,繁體中文-非偏好,UF, 自然景觀 泥坑(沼澤地)(,U,繁體中文-非偏好,UF, 草澤地 (,U,繁體中文-非偏好,UF,U,U)濕地 Non-preferred terms 沼地 (草澤) (.U,繁體中文-非偏好,UF,U,U) under the same facet 沼原 (,U,繁體中文-非偏好,UF,,) and hierarchy **Terms in** subject id qualifier **Chinese scope notes Rules** Chinese Chinese marshes 沒有林木生長的低窪溼地,通常雜草、香蒲及燈 沼地/zhao di 300008903 preferred 心草叢生。 (non-preferred) 草澤 term Chinese slews 充滿泥濘或長滿蘆葦的水池、池塘、小湖泊、水 沼地/zhao di 300343521 preferred 沼澤地 窪或水灣地區。 (non-preferred) term





From Chinese to English/ Localization

• Suggestion: Topics on Festivals of Taiwanese Indigenous People

New Terms Suggestion: Festivals of Taiwanese Indigenous People

- Taiwan is a multi-cultural island enriched by Chinese and local indigenous cultures; also homeland of the Austroneasian people.
- Since AAT has terms about Chinese Festivals (8 records), as initial idea we suggest 16 candidate concepts for the Festivals of Taiwanese indigenous people, as 16 different tribes of indigenous people are existed in Taiwan.



Tribes in Taiw	/an	Festivals	
Amis	阿美族	Malalikit	豐年祭
		(Harvest Festival)	
Atayal	泰雅族	Ps I kautas	祖靈祭
-		(Ancestor Worship Festival)	
Paiwan	排灣族	Maleveq	五年祭
_		(Five-Year Worship)	A 1 6
Bunun	布農族	Ma-naq-tainga	射耳祭
		(Shoot The Ear Festival)	VI
Puyuma	卑南族	Mangamangayau	猴祭
		(Monkey Ritual)	
Rukai	魯凱族	Tsatsapipianu	小米收穫祭
	ATT L	(Millet Harvest Ritual)	
Tsou	鄒族	Mayasvi	戰祭
		(Victory Ceremony)	本言な
Saisiat	賽夏族	Pas-taai	矮靈祭
_	いた (ナントト	(Dwarf Spirit Ceremony)	The for the start
Тао	達悟族	Do rayon	飛魚祭
These	᠇᠋ᡕᢣᡔ	(Flying Fish Ceremony)	
Thao	邵族	Lus'an (Animist Worship)	祖靈祭
Kavalan	葛瑪蘭族	Palilin (Ancestor Rite)	歲末祭祖
Truku	太魯閣族	Mgay Bari (Ancestor Rite)	感恩祭
Sakizaya	撒奇萊雅族	Palamal	火神祭
		(Worship of the Fire God)	
Seediq	賽德克族	Kmetui macu	小米播種祭
		(Millet Sowing Festival)	
Hla'alua	拉阿魯哇族	Takiaru (Sacred Shell Rite)	貝神祭
Kanakanavu	卡那卡那富族	Mikong (Appreciation for	米貢祭
		Harvest Ceremony)	

16 Festivals of Taiwanese indigenous people



Malalikit festival of Amis



Source:

http://www.apc.gov.tw/portal/index.html

Applications & Future Plans • Linked Taiwan Artists (LTA) • Union Catalog LOD Program (UC-LOD) • DigiMuse: Digital Indigenous Museum with AAT • Early Chinese Periodicals Online (ECPO)

AAT-Taiwan application program 1

Linked Taiwan Artists (LTA) project with AAT



Project Purpose

- Promoted by the Getty Research Institute in recent years, development and application of Linked Open Data (LOD) are becoming central issue in research. Since 2014, AAT, TGN and ULAN are published as LOD and provide worldwide users an important reference in the study field of KOS.
- Since LOD is the current trend in the research of digital humanities, AAT-Taiwan initiated the project "Linked Taiwan Artists (LTA)" based on the database of previous project "Starting out from 23.5° N: Chen Cheng-Po" and KOS structured by Getty.

Ontology of LTA

LTA-ontology: describing datasets of Taiwanese artists in semantic models:

- 30 classes are organized under 6 main categories: date, person, event, place, object, organization.
- LTA datasets reuse 57 diverse properties from 7 RDF vocabularies (gvp, schema, cidoc-crm, rdfs, skos, dc, dbpprop) to structure the triples.



Properties (57)

[ObjectProperties (39)]

crm:P2_has_type ; crm:P11_had_participant ; crm:P12_was_present_at ; crm:P12i_occurred_in_the_presence_of ; crm:P14_carried_out_by ; crm:P15_was_influenced_by ; crm:P45_consists_of ; crm:P55_has_current_location ; crm:P62_depicts; crm:P70_is_documented_in; crm:P71_is_listed_in; crm:P108_has_produced ; crm:P130_features_are_also_found_on; gvp:ulan1313_partner_in; gvp:ulan1501_sibling_of ; gvp:ulan1511_child_of ; gvp:ulan1512_parent_of ; gvp:ulan1541_spouse_of ; gvp:ulan1550_relative_by_marriage ; gvp:ulan2550_friend_of ; gvp:ulan2828_student_at ; schema.org: homeLocation; schema.org: nationality; schema.org:about; schema.org:attendee ; schema.org:birthPlace ; schema.org:foundingLocation ; schema.org:genre ; schema.org:hasMap ; schema.org:location ; schema.org:organizer ; schema.org:recordedIn ; schema.org:replacee ; schema.org:replacer ; schema.org:subEvent ; schema.org:worksFor [DataProperties (14)]

crm:P1_is_identified_by; crm:P3_has_note; dbpprop:years; dc:format; schema.org:award; schema.org:birthDate; schema.org:dateCreated; schema.org:deathDate; schema.org:dissolutionDate; schema.org:endDate; schema.org:foundingDate; schema.org:startDate; schema.org: latitude; schema.org:longitude

[AnnotationProperties (4)]

rdfs:label ; rdfs:comment ; rdfs:seeAlso ; skos:altLabel

Properties of LTA ontology

LTA ontology with Getty vocabularies

- GVP vocabularies (especially properties from gvp:ulan) are adopted to give correct description of personal relationships between artists.
- rdfs:seeAlso: property linking LTA datasets with Getty terms of AAT, ULAN, GVP



Triple Examples of Objects in LTA

	Subject	Predicate	Object	Example				
1	Object	crm:P1_is_identified_by	ID	Stream - crm:P1_is_identified_by - A17				
2		rdfs:label	(文字) string literal	Stream - rdfs:label - Steam				
3		skos:altLabel	(文字) string literal	Stream – skos:alLabel – 清流				
4		crm:P3_has_note	(文字) string literal	Stream - crm:P3_has_note -In early 1929, Chen Cheng-po took another trip(畫作簡介,後略)				
5		rdfs:seeAlso	URI	Stream – rdfs:seeAlso - http://chenchengpo.asdc.sinica.edu.tw/treasures?id=A17&pid=0701				
6		crm:P2_has_type	AATDisplayi ngItem	Stream - crm:P2_has_type - http://aat.teldap.tw/AATFullDisplay/300015636				
7	ObjectEvent	schema.org:startDate	Time	"Chen Cheng-po painting Stream "-schema.org:startDate - ?				
8		schema.org:endDate		"Chen Cheng-po painting Stream "- schema.org:endDate -1929				
9		crm:P14_carried_out_by	Person	"Chen Cheng-po painting Stream "– crm:P14_carried_out_by – Chen Cheng-po				
10		crm:P108_has_produced	Object	"Chen Cheng-po painting Stream "–crm:P108_has_produced – Stream				
11		crm:P15_was_influenced _by	Object	"Chen Cheng-po painting SpringAtWestLake_2"- crm:P15_was_influenced_by – Hangzhou-SB19 : 34				
* coli	umn in orange	refers to predicate shared by dif	ferent types of	things. Major ontologies in use: CIDOC-CRM(12) schema.org (4)				

	Subject	Predicate	Object	Example
1	Event_ Exhibition	crm:P1_is_identified_by	ID	The1 st TaiwanArtExhibition - crm:P1_is_identified_by - EA001
2		rdfs:label	string literal	The1 st TaiwanArtExhibition - rdfs:label - The1 st TaiwanArtExhibition
3		skos: alt Label	string literal	The1 st TaiwanArtExhibition – skos;altLabel - The 1 st Taiten
4		crm:P3_has_note	string literal	The1 st TaiwanArtExhibition – crm:P3_has_note - The new art form gradually took shape after the launch of the Taiwan Art Exhibition (Taiten) in 1927(展覽簡介,後略)
5		rdfs:seeAlso	URI	Chicago World's Fair - rdfs:seeAlso - http://en.wikipedia.org/wiki/Century_of_Progress#/media/File: Chicago_world%27s_fair,_a_century_of_progress,_expo_poster, _1933,_2.jpg
6		schema.org:startDate	Time	The1 st TaiwanArtExhibition – schema.org:startDate – 1927-10-27
7		schema.org:endDate	Time	The1 st TaiwanArtExhibition – schema.org:startDate – 1927-11-06
8		schema.org:location	Place	The1 st TaiwanArtExhibition – schema.org:location – Kabayama Elementary School Taipei
9		crm:P2_has_type	Data	The1 st TaiwanArtExhibition - crm:P2_has_type - http://aat.teldap.tw/AATFullDisplay/300033799 (AAT oil painting)
10		crm:P11_had_participant	Person	The1 st TaiwanArtExhibition – crm:P3_has_note – Chen Cheng-po

Characteristics of LTA

Linked Taiwan Artist (LTA): an open datasets of Taiwanese artists in the Japanese colonial period (1895-1945). LTA has following characteristics:

- 1. Open Data: allow others to reuse the data
- 2. Supporting Faceted Search/ Semantic Browsers
- **3. Expose the Context of Research Materials**
- 4. Integrate Heterogeneous Data into a Unified Linked







Suppor	ting Fac	eted Search/	Semantic Browsers
		Taiwan Artists	PLACE TIME
Search Chen Cheng-po Chen Jin Chen Zhi-qi Guo Xue-hu He De-lai	relationsh	h results show differe ip between Chen Che hun by system's indu	eng-po and
□Huang Tu-shui			► More
🗆 Itaru Tanabe			
□Kinichiro Ishikawa	Tal		Chen Cheng-po - crm:P11_had_participant - 9 th Taiwan Art Exhibition
□Li Mei-shu	Title	Taiwan Art Exhibition_9th	
□Li Shih-chiao ■Liao Chi-chun	Participant	Chen Cheng-po 、 Liao Chi-chun	Liao Chi-chun - crm:P11_had_participant - 9th Taiwan Art Exhibiton
Liao Chi-chun			
Ni Chiang-huai			
□Pan Yu-liang			22 AM
□Pu Tian-sheng			Chi-Yang Western Painting Society
Saburosuke Okada	Title	Chih-Yang Western Painting Society	+ = +
□Sabulosuke Okada	Member	Chen Cheng-po × Liao Chi-chun	, ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
Detsuomi Tateishi	Weniber	Cherroneng-po · Elao chi chun	
□Toho Shiotsuki			
□Wang Bai -yuan			Member 19
□Wang Ji-yuan			
□Yan Shui-long			
□Yang San-lang	Title	Tokyo School of Fine Arts	
□Yu Jian-hua	Alumnus	Chen Cheng-po 、Liao Chi-chun	
□Zhang Qiu-hai	Participant	Chen Cheng-po 、Liao Chi-chun	and the second
□Zhang Yi-xiong	earch		Tokyo School of Fine Arts 東京美術學校

Diverse interactive relationship or potential relevance between artists can be revealed and inferred at the same time, such as members or alumni at the same organization.





AAT-Taiwan application program 2

Union Catalog-LOD program with AAT



Project purposes

- The Taiwan Digital Archives Union Catalogue (UC) (<u>http://catalog.digitalarchives.tw/</u>) : an online portal for the digital preservation of cultural heritages with over 5 million Dublin Core(DC)-based metadata records.
- The project aims to reconstruct the whole system as semantic LOD datasets and link with Getty vocabularies. A pilot study has been conducted with 850,000 data from 72 collections across 16 categories licensed under attribution of Common Creatives.
- The UC-LOD datasets use a variety of RDF vocabularies and ontologies as fellows:
 - DC: vocabulary of basic metadata terms (15 properties)
 - dcterms: terms for describing advanced metadata (46 properties)
 - DWC: terms for specifying biological metadata (10 properties)
 - Schema: describing measurement or video resources (15 properties)
 - WGS84: for geographic info (3 properties)
 - RDFS: terms for directing title of resource (1 property)



Regenerate and normalize original xml data

- UC as a long-term project, nearly 100 organizations are involved. Metadata and contents are set up by different staff. However, incoherence in metadata fields and values between sub-projects are always problem. Metadata of all projects should be checked and cleaned at first.
- Check metadata structures in original xml files; normalized by UC-LOD Attribute Field Normalizer, a tool programed by team AAT-Taiwan.

- <metadesc license="CC3.0:BY-NC-ND"></metadesc>		- <metadesc license="CC3.0:BY-NC-ND"></metadesc>
<title>拉丁屬名:Thlaspida</th><th></th><th><Title field="拉丁屬名">Thlaspida</title>		
<title>拉丁種名:formosae</eu</th><th>- CD Am/bute Reld Normalizer</th><th>A statistic s</th></tr><tr><th><Subjec>拉丁門名:ARTHROPODA</S</th><th>UC-LOD Attribute Field Normalizer - 1.0.6</th><th><Subject field="拉丁門名">RTHROPODA</Subject></th></tr><tr><th><Subjec>中文門名:新技動物門</Subjec</th><th>GG-COD Attribute Field Hormanzer - 1.0.0</th><th> <Subject field="中文門名"> # 鼓動物門</Subject></th></tr><tr><th></th><th>s directory: 「韓」時時期18月1日 会」)加以韓」時代由東京市会会設計的時代台湾市高品牌構成的objy 1 Choose Fo</th><th></th></tr><tr><th><Subject>中文辦名:旨器辦</Subject></th><th>directory: C'Users'use/Desktoo'Biology 02 Insect NTU</th><th><Subject field="中文鋼名"> 提a鋼</th></tr><tr><th> Subject 3拉丁亞明社 - Prerygota </th><th></th><th>Subject field="拉丁亞鋼名"。Pterygota</Subject></th></tr><tr><th><Subject>中文亞鍋名:有翅亞鍋</Subject</th><th></th><th> <Subject field="中文亞綱名">有翅亞綱 </th></tr><tr><th><Subject>拉丁目名:Cpleoptera</Subj</th><th></th><th><Subject field="拉丁目名">Goleoptera</Subject></th></tr><tr><th><Subject>中文首名:新期目</Subject></th><th>Choose fields to modify:</th><th><Subject field="中文目名">#週目</Subject></th></tr><tr><th></th><th>inte 🖉 Creator 🖉 Subject 🖉 Description 🖉 Publisher 🖉 Contributor 🖉 Date 🖉 I</th><th></th></tr><tr><th></th><th>Format 🔄 Mentifier 🔄 Source 🔄 Language 🔄 Relation 🔄 Coverage 📈 Rights</th><th><Subject field="中文亞目名" 多食亞目</Subject></th></tr><tr><th><<u>Subjec</u>>拉丁料名:Chrysomelidae<//</th><th>(R.9. 🛁</th><th><Subject field="拉丁料名">•hrysomelidae</Subject></th></tr><tr><th><Subject>中文科名:全<mark>花盡科</Subject></mark></th><th>(i) The normatizing is finished</th><th> <Subject field="中文科者"><注范蟲科</Subject > </th></tr><tr><th><Descri<mark>tion>食性:植食性</Description</mark></th><th>The formalizing is initialized</th><th><Description field="食性"># 食性</Description></th></tr><tr><th><Descri<mark>tion。吕否为好</mark>约動物:否(No)<,</th><th>1025</th><th>start Decemption fuld "是百為鼎有動物">否(No)</Description></th></tr><tr><th><Description>模式類型:群模式標本(syn</th><th></th><th><Description field="視式識型"> 新根式標本(syntype)</Description></th></tr><tr><th><Description>標本狀況:良好(good)保存 znzz</th><th>· XI M TX-1 - COORTINGTON</th><th> <Description field="標本狀況">良好(good)保存方法:針插標本</Description></th></tr><tr><th><Date>採集期間-起:1919/2/25</Date></th><th></th><th><Date field="孫集期間-起">1919/2/25</Date></th></tr><tr><th><Date>採集期間-迄:1919/3/27</Date></th><th></th><th><Date field="孫集期間-迄">1919/3/27</Date></th></tr><tr><th><Type>昆蟲、備雄規模本</Type>
<Format>媒體種類:數位影像檔</Format> =</th><th></th><th><Type>品品、?類類様本</Type></th></tr><tr><th></th><th></th><th><Format field="課證種類">數位影像檔</Format></th></tr><tr><th><Coverage>原始採集資訊: Taiwan Taito</th><th>Normalizing incoherent</th><th><Coverage field="原始採集資訊"> Taiwan Taito 25. II27. III. 1919 Col</th></tr><tr><th><Coverage>中文國名:台灣</Coverage>
<Coverage>英文國名:Tawian</Coverage</th><th></th><th><Coverage field="中文國名">台灣</Coverage>
<Coverage field="英文國名">Tawian</Coverage></th></tr><tr><th><Coverage>英文國有, Tawlan (Coverage)
<Coverage>英文第一行政分置: Taitung H</th><th></th><th></th></tr><tr><th><Coverage>央文第一行政分置: faituing in</th><th>original metadata by UC</th><th><Coverage field="中文第一行政分置">台東數</Coverage></th></tr><tr><th><Coverage>英文第二行政分置:百來版</Coverage>英文第二行政分置:Taitung </</th><th></th><th><Coverage field="英文第二行政分置">Taitung </Coverage></th></tr><tr><th><Coverage>英文第二行政分域:台東市</Co</th><th>Etald Name altern</th><th><Coverage field="中文第二行政分響">faiturig </coverage></th></tr><tr><th><Rights>行政院總署委員會應著試驗所</Right</th><th>Field Normalizer</th><th><Rights>行政院農業委員會農業試驗所</Rights></th></tr><tr><th><Rights>(http://www.tari.gov.tw/)</ki</th><th></th><th><Rights>(http://www.tari.gov.tw/)</Rights></th></tr><tr><th></r> </r></th><th>Auro.</th><th></mailes/mailes/</th></tr><tr><th></DACatalog></th><th></th><th></DACatalog></th></tr><tr><th>Sy Discussing P</th><th></th><th></th></tr><tr><th></th><th></th><th></th></tr></tbody></table></title>		

Parse normalized data and produce triple-structures

- Original UC metadata fields are structured after Dublin Core, to apply it as triplestructured semantic data for LOD, recheck and proof their accuracy or make an adjustment by adopting other semantic vocabularies are necessary.
- For this purpose, we develop another tool UC-LOD XML Parser, to
 - parse and convert normalized xml data into json files;
 - import json files in open refine to produce triple-structure.

Settings filepath: C:\Users\user\Desktop\)	Refresh	Reset All Remove All	Show as: r	ows records Show:	5 10 25 50 row	s « firs	t « previous 1 - 50 next » la
	聯合目錄LOD計畫\b. Parsing XML\p	Choose File	×p	change		_ -p	▼s	▼t ▼I	
Rule filepath: C:\Users\user\Desktop\	聯合目錄LOD計畫\b. Parsing XML\cd	Choose File	59 choices Sort by: na		☆ 디 1. ☆ 디 2.	Title-學名 Subject-界	4654125		Bracon adoxophyesi Minami 動物界
KMLs direcotry: 物類_文書類_文物類_碑	拓類18項計畫_XML檔_履彥\農業試驗	Choose File	Contributor-採集者 171 Contributor-採集者英文		☆ 덕 3.	Subject-界 (英文)	4654125		Animalia
			名 3423		公 5 4.		4654125		節肢動物門
Dutput file name: C:\Users\user\Desktop\E	Biologhy_TARI-2_RAW.json	Choose File	Contributor-標本資訊訂	正者英	1 13 9 5.	Subject-門(英文)	4654125		Arthropoda
100%	1	Start	文名 91		☆ 듸 6.	Subject #	4654125		昆蟲綱
100%		Start	Coverage 州省 4019		😒 🗐 7.	Subject-鋼(英文)	4654125		Insecta
	訊息	x	Coverage-州省(英文)	4019	☆ 다 8.	Subject-亞編	4654125		有樹亞綱
arse and write 5982788.xml			Coverage-次縣級行政	g 1184	1 😒 🗐 9.		4654125		Pterygota
arse and write 5982789.xml	Parsing finished.		Coverage 次縣級行政	直(英		Subject-目	4654125		態想目
arse and write 5982790.xml			文) 1183			Subject-目(英文)	4654125		Hymenoptera
arse and write 5982791.xml	隆定		Coverage/洋洲 4399			. Subject-西昌	4854125		編腰亞昌
rocess finished) <u>)</u>)	Coverage-國家 1			. Subject-亞目 (英文)	4654125		Apocrita
	C		Coverage-國家(英文)			. Subject-科	4654125		小鶴蜂科
	-		Coverage 採集地 4270)		. Subject-科(其文)	4654125		Braconidae
			Coverage-採集地(英			Subject-超科	4654125		姬蜂總科
			文) 4393			Subject-超科(英文)	4654125		Ichneumonoidea
			Coverage 1 2032		1 🕹 🖾 18	Subject-苏彩	4654125		當種百彩

Reconcile metadata values by command tester

- In open refine, we check accuracy of metadata or make data cleaning if necessary.
 Hence, UC-LOD command tester is designed to clean and reconcile metadata values by giving them with standardized rules in tester.
- UC-LOD command tester: In general
 - 11 different rules are programed to satisfy different needs
 - Rules can be used in an additive way. Basic rule norms are included to make convenience to general users.

	🛃 UC-LOD Command Tester					11	L different rules in Comman	d lester	
		LIC LOD Command	Tester - 1.1.6			Rules	Rule Example@	Before <i>e</i>	After
條斑螟小繭蜂	UC-LOD Command Tester - 1.1.6			條斑螟小繭蜂 姬蔗龜	小繭蜂 between↔	<between <br="" after="(">before=")"></between> ↔	(Hawaii)↔	Hawaii₄	
如蔗龜;小蔗龜 稻蝨、黑腳煙仔 稻混蟲,稻葉蟲 Zigzag striped		characters>chinese			小蔗龜	contains ₽	<contains <br="" all="true">opposite="false">A B</contains> ↔	ABC* ²	AB⊷
		ľ	稻蝨 黑腳煙仔	Characters₽	<characters>chinese</characters> φ	臺灣? Taiwan?	臺灣 [。] 足球 soccer ^{。)}		
leafhopper					稻泥蟲 稻葉蟲			足球 soccer+ [,] football+ [,]	
		InputType: Text v	t			Length₽	<length larger="true">4</length> ↓ ³	Getty≓ GRI€	Getty₽
		Input Area 條斑螟小蘭蜂	Output Area 條料螟小蘭蜂			Insert	<insert position="begin">Note: </insert> + ⁰	Good₽	Note:Good₽
		姬蔗龜:小蔗龜 稻蝨、黑腳煙仔	姬蔗龜 小蔗龜			remove* ²	<remove><i> </i></remove> +	Thlaspid a +2	Thlaspida+2
		稻泥蟲, 稻葉蟲 Zigzag striped leafhopper	稻蝨 黒脚煙仔 怒泥典			remove_be tween₽	< <u>remove_between</u> after="(" before=")" include="false"> <u remove_between>+ ³	Hello,(test) how are you⇔	Hello,() how are you↔
						replace	<replace new=" > ">,</replace> +	Tigre₽	Tiger↔
	Te	sting rules for d	-	; k	ру	Split ₽	<split> \ </split> ₽	TAIWAN ► USA+ ²	TAIWAN⊷ USA⊷
		UC-LOD comm	hand tester			Value₽	<value>fixed words</value> ↔	台北↩ 高雄↩	fixed words↩ fixed words↩

11 different rules in Command Tester
Establish correct rule files

- Establish and rewrite rules for each subproject to produce correct triple structures for UC-LOD project
- Give accurate language value, type value, rules for data cleaning and predicates in files to structure metadata triples



Prove accuracy of rule files and import in triplestore

- Generate reconciled triple structure by open refine and prove its accuracy
- Import datasets into UC-LOD triple store



UC-LOD SPARQL endpoint- platform for query and download

Taiwan 典藏臺灣 Digitalarchives

- Data visualization: UC-LOD provides
 SPARQL endpoint with user-friendly interface to query, search and download organ the wished LOD data.
- Functions provided in SPARQL endpoint:
 - offer diverse selectable query models
 - convenient insertion of Prefix/ Predicate
 - free downloads of data in json, XML, TTL or CSV file







Linking UC-LOD with AAT

In this project we plan to link UC-LOD datasets with AAT vocabularies to increase their discoverability and extend applicability through structured LOD model! An example from UC metadata: Handloom

	Subject		Туре		Coverage		Holding location
value	紡織機		實體物件		卑南 (臺東)		中央研究院
Getty vocab.	AAT 3000 handloom		AAT 300117 artifacts (ol genre)		TGN 7491221 Taidong (county)	ULAN 500309388 Academia Sinica
Click the A icon to v Semantic View (ISC ID: 300023505 A handlooms (it (hierarchy nam Note: Any of variou: Terms: handlooms (it handlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v Andlooms (it (bierarchy nam Note: Any of variou: Click the A icon to v	N. JSONLD. RDF. N3/Turtle. N- oms (textile tools), <textile fab<="" td=""><td>Click the & icon to Semaric View (3) Click the & icon to Semaric View (3) TD: 300117127 artifacts (obj artifacts (obj a</td><td>& Architecture Thesi I Record Display</td><td>Click the koon to Semartic View (JS) To: Tailong (coor Coordinater: Lat: 22 53 01 N of Coor Coordinater: Lat: 20 54 E</td><td>Thesaurus of Geographic Names' ecord Display «Inclose Page week the Nierarchy. OK JSCNAD. RDE. NOTurtie. N.Treden) atv) process minutes Lat: 22.8836. decimal depreses depress minutes Lat: 22.8836. decimal depreses erred NAv, Chinese (transitiented Rivin without tones),(b) (Chinese (transitiented Rivin without tones),(b) (Chinese (transitiented Rivin without tones),(b) (Chinese (transitiented Rivin) without tones),(b) (Chinese (transitiented Rivin) (transitiented Rivin) (transitiented Rivin) (transitiented Rivin) (transitiented Rivin) (transitiented Rivin)</td><td>Click t Sema ID: 5 A Name Ac Natio Ta Roles rep</td><td>Union List of Artist Names® Online Full Record Display Providus Page te icon to view the hierarchy. Inter View (USON, USONLD, RDE, N3/Turtle, N-Toples) 000309368 cademia Sinica (Taiwanese repository, contemporary) Esi addemia Sinica (preferred,U,display) natifities: Wanese (preferred)</td></textile>	Click the & icon to Semaric View (3) Click the & icon to Semaric View (3) TD: 300117127 artifacts (obj artifacts (obj a	& Architecture Thesi I Record Display	Click the koon to Semartic View (JS) To: Tailong (coor Coordinater: Lat: 22 53 01 N of Coor Coordinater: Lat: 20 54 E	Thesaurus of Geographic Names' ecord Display «Inclose Page week the Nierarchy. OK JSCNAD. RDE. NOTurtie. N.Treden) atv) process minutes Lat: 22.8836. decimal depreses depress minutes Lat: 22.8836. decimal depreses erred NAv, Chinese (transitiented Rivin without tones),(b) (Chinese (transitiented Rivin without tones),(b) (Chinese (transitiented Rivin without tones),(b) (Chinese (transitiented Rivin) without tones),(b) (Chinese (transitiented Rivin) (transitiented Rivin) (transitiented Rivin) (transitiented Rivin) (transitiented Rivin) (transitiented Rivin)	Click t Sema ID: 5 A Name Ac Natio Ta Roles rep	Union List of Artist Names® Online Full Record Display Providus Page te icon to view the hierarchy. Inter View (USON, USONLD, RDE, N3/Turtle, N-Toples) 000309368 cademia Sinica (Taiwanese repository, contemporary) Esi addemia Sinica (preferred,U,display) natifities: Wanese (preferred)

AAT-Taiwan application program 3

DigiMuse: Digital Indigenous Museum with AAT



Project purposes

- "Digital Indigenous Museum", a Pilot project of DigiMuse program organized by ASCDC
- Interactive platform for digital preservation and reuse of digital cultural resources in Taiwan for collection, research, demonstration and authorization for users.
- Museums, content holders, and the general public can upload its resources for integration and reproduction, continue knowledge sharing in public.
- As initial stage we focus on cultures of Taiwanese indigenous people with ca. 300 selected items:
 - Parsing metadata from selected fields (objects, materials) and mapping to AAT
 - Displaying by knowledge graph







Project purposes

ECPO Keywords

Arts and Crafts

Painting

工藝美術

Exhibition

繪畫

展覽

- Early Chinese Periodicals Online (ECPO): a joint program between Academia Sinica and Centre for Transcultural Studies of Heidelberg University (HCTS).
- In ECPO system, 230 kinds of periodicals, published between 1904 and 1937, with focus on women, gender issues, arts, literature or entertainment are digitalized. 879 keywords are set up to structure textual and image topics in ECPO.
- Current study: mapping 105 keywords from 238 images collected in "The Crystal", a boulevard magazine issued between 1919 and 1940 in Shanghai, with AAT or TGN concepts.

facet

Tagging in AAT concepts

paintings(visual works)

AAT 300033618

AAT 300250482

AAT 300054766

exhibitions(events)

industrial arts



Mapping Results of ECPO Keywords to AAT/ TGN

Facet	Hierarchy	Keywords	Term Frequency
Associated Concepts Facet (Type 19; Value 111)	Associated Concepts	19	111
Agents Facet	People	11	61
(Type 15; Value 72)	Organizations	2	3
	Living Organisms	2	8
Activities Facet	Disciplines	9	55
(Type 25; Value 116)	Functions	3	21
	Events	7	22
	Physical and Mental Activities	2	3
	Processes and Techniques	4	15
Materials Facet (Type 9; Value 15)	Materials	9	15
Objects Facet	Built Environment	6	8
(Type 31; Value 89)	Furnishings and Equipment	10	22
	Object Genres	5	37
	Object Groupings and Systems	1	2
	Visual and Verbal Communication	9	20
Geographic Names (TGN)	Continent	1	1
(Type 6; Value 12)	Nation	5	11
Total		105	415



Lin

LIDE:

9 metally



以視覺化呈現資料之間的脈絡(巨觀與微觀) 並提供進一步的探索

The Open Graph Viz Platform

Gephi is the leading visualization and exploration software for all kinds of graphs and networks. Gephi is open-source and free.



GTAA 16

Summer of



Thanks for your listening



The Getty Research Institute

菜牛, 建菜京引典

