Academia Sinica Center for Digital Cultures
中央研究院數位文化中心

Framing Digital Research Environment in Academia Sinica of Taiwan

Sophy Shu-Jiun Chen

Executive Secretary, Academia Sinica
Center for Digital Cultures (ASCDC), Taiwan (R.O.C.)

Assistant Research Fellow, Institute of History and Philology, Academia Sinica

August 23, 2018

Brief History of Taiwan Digital Archives

- **1998-2001 pilot phase (three pilot projects)**
  - Digital Museum Project
  - National Archive Digitization Program
  - International Digital Library Cooperation Plan
- **2002-2007 phase one (two national programs)**
  - National Digital Archives Program (NDAP)2002-2007
  - National Science and Technology Program for e-learning (ELNP)2003-2007
- **2008-2012 phase two (one integrated program)**
  - Taiwan e-Learning and Digital Archives Program (TELDAP)
- **2013- main project (Sustainable Achievements)**
  - **Academia Sinica Center for Digital Cultures**
    - Phase 1: 2013-2015
    - Phase 2 : 2016-present

Footage of Taiwanese classical film in 1950s
(Chinese Taipei Film Archive)

Digitization Procedures Guideline Series

Digitization Procedures Guideline series were published. These publications are divided into “common concept” and “selected objects” based on two different perspectives regarding collection as a whole and a single type. A complete and comprehensive version can be accessed online according to individual needs. The purpose of these Guidelines is to provide proactive and practical references and amplify the efficiency and quality of digitization.

Achievement Statistics (Taiwan Digital Archives)

100 +
participated Institutions

Libraries/Archives/Museums/ Academic Institutions/Private Institutions/Private Collectors…etc

Over 5 million
digitized objects

Anthropology/Archaeology/ Art History/Biodiversity/Rare Books/Archives/Maps…etc
**selected digital collections of Academia Sinica*

750 +
websites & databases

Digital Resources for Sinology Studies in Taiwan / Reading Digital Atlas / The Painter Chen Cheng-Po…etc

Goals of Academia Sinica Center for Digital Cultures (ASCDC)

- Maintain and preserve Taiwan Digital Archives
- Develop a digital environment, including platforms of digital humanities, digital curating, digital publishing and core technologies(such as text analysis, image annotation, linked open data…etc.)
- Enhance further on digital humanities, digital cultures and creative applications.
- To facilitate the advancement of digital technology in research, education, culture and industry development to create an integrative and innovative content environment.

Digital Resources for Sinology Studies in Taiwan

臺灣漢學數位典藏資源導覽

中文版 English http://sinology.ascdc.sinica.edu.tw/index_en.html

HOME | ARCHIVES | FULLTEXT | ANTHROPOLOGY | RARE BOOKS | RUBBINGS | ART & SOUNDS | MAPS | OTHERS | 2016 AAS | 2017 AAS

Academia Sinica Digital Archives Maps Search System

Center for GIS, RCHSS, Academia Sinica
http://map.rchss.sinica.edu.tw

This website provides access to 12 databases, including: Maps Collected in the Inst. of Modern History, Maps Collected in the Inst. of History and Philology, Replica Topographic Maps of Mainland China Collected in the Japanese Science Academy, Topographic Maps of Mainland China Collected in the Ministry of National Defense, Land Surveyed Topographic Maps of China Provinces Collected in the Department of Land Administration, Thematic Maps Collected in the Department of Land Administration, Historical Cadastral Maps Collected in the National Land Surveying and Mapping Center, etc.

Reading Digital Atlas

Academia Sinica Digital Center
http://digitalatlas.asdc.sinica.edu.tw

The bi-lingual website contains 157 pre-1900 historical Chinese maps from the Library of Congress. The maps selected in this website are either of academic significance, made with exquisite craftsmanship or rare to be seen among the collection institutes. They are organized into 9 categories: (1) general maps (2) local administrative maps (3) flood control and embankment maps (4) military and traffics maps (5) city maps (6) political and boundary maps (7) maps of temples, palaces and royal gardens (8) maps of social and economic conditions (9) mixed-theme maps. All maps are accompanied with detailed metadata and annotations.

Chinese Civilization in Time and Space

Academia Sinica
http://ccts.sinica.edu.tw

Based on Tan Qixiang's The Historical Atlas of China, the system provides fundamental base maps of China over the past 2000 years, from the ancient time to Qing dynasty, supplemented by various historical atlas and remote sensing imagery.

GIS of the Literature and Historical Documents of Su Shi

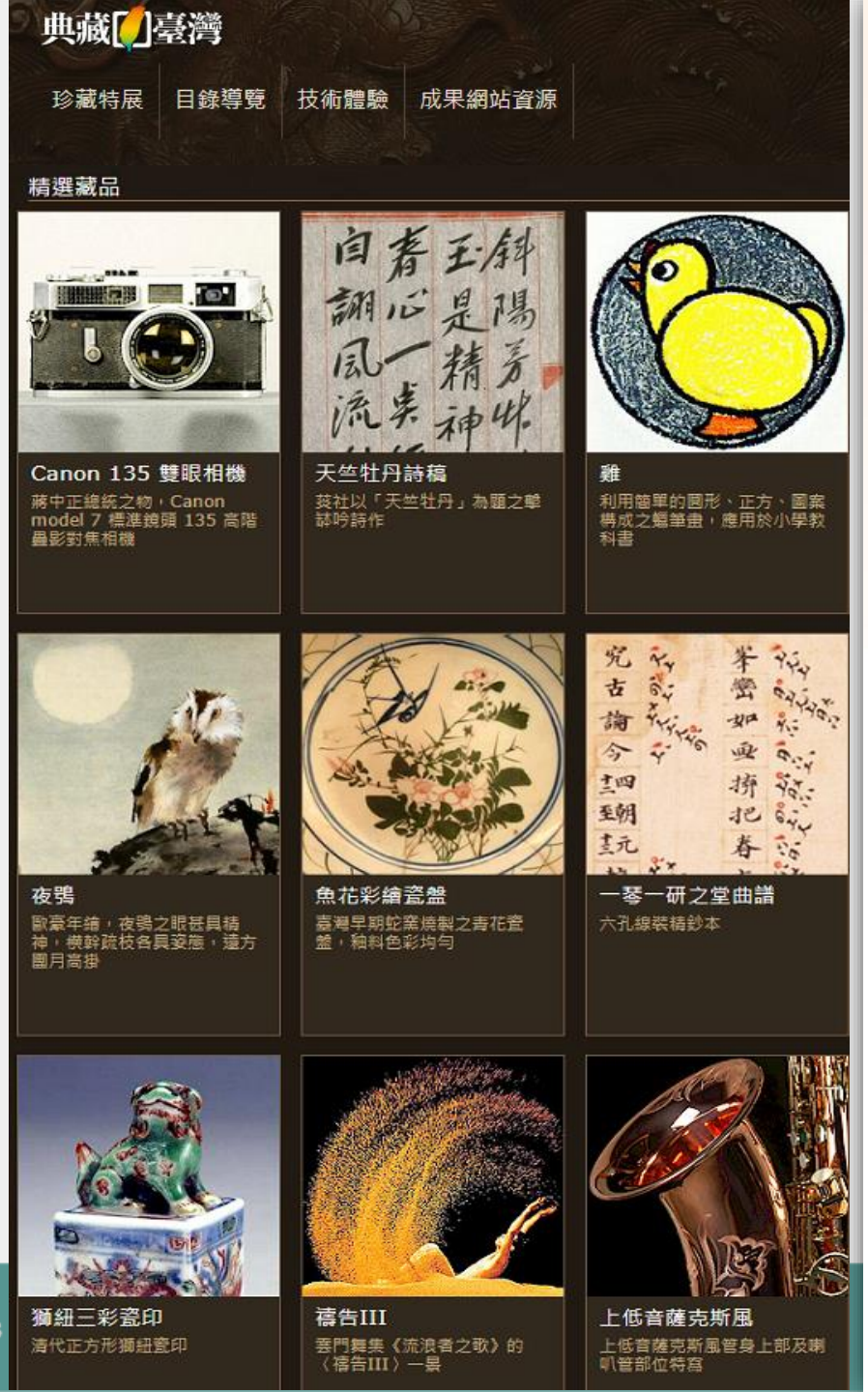
Yuan Ze University
http://cls.hs.yzu.edu.tw/su_shi/

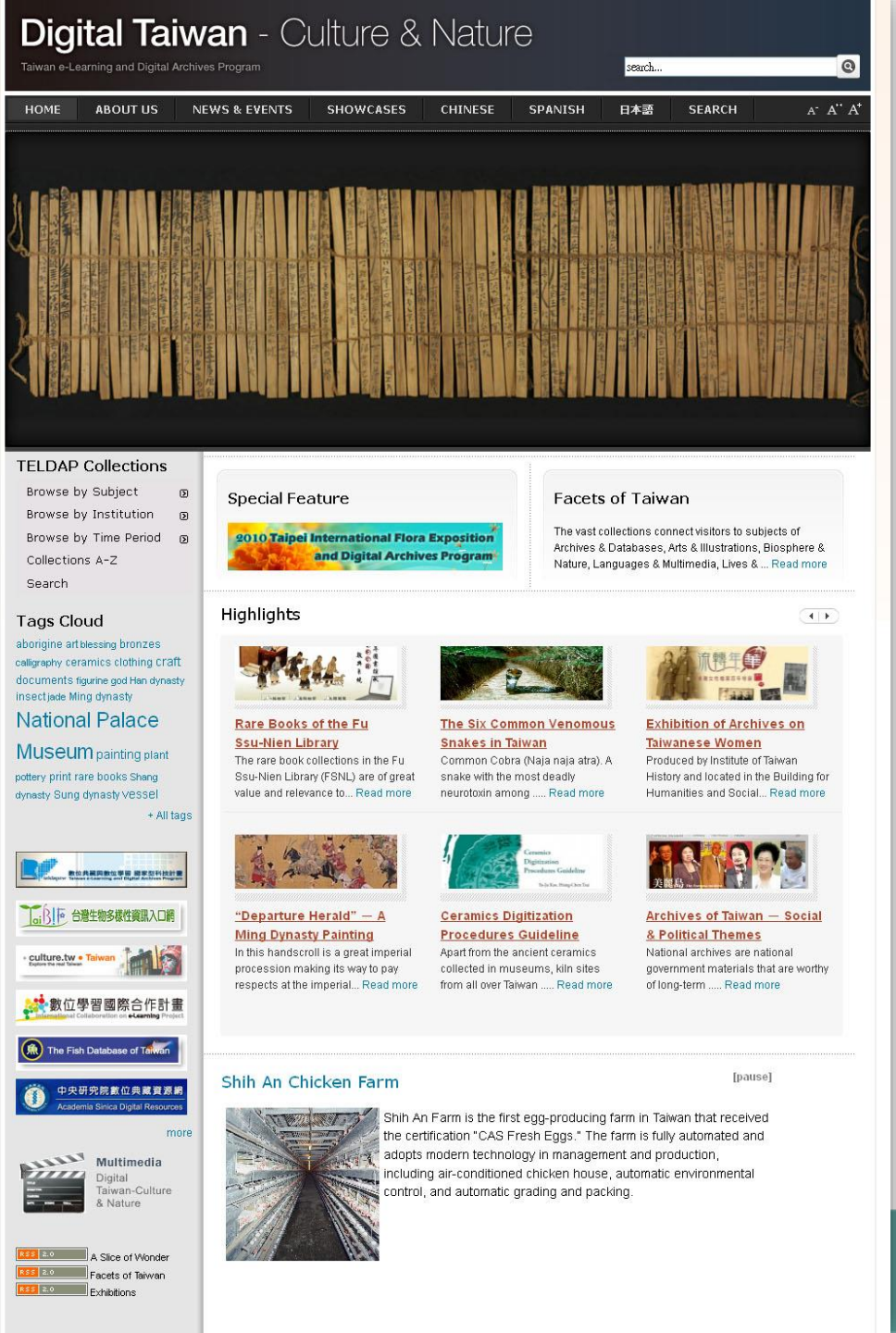
Based on Tan Qixiang's atlas of Song dynasty, with maps in local gazetteers, admisnistrative maps, and the literary works and poems written by Su Shi, the system builds a Su Shi map covering people, things, times and places, which serves as an

GIS of the Distribution of Authors and Works of Tang Shi and Song Ci

Yuan Ze University
http://cls.hs.yzu.edu.tw/TSPL/

The system uses GIS and other information tools to build a distribution and transition map of the authors and works of Tang Shi and Song Ci. The distribution data of works not only correlate closely with the authors, but also include places related to the works





The Portal of Taiwan Digital Archives

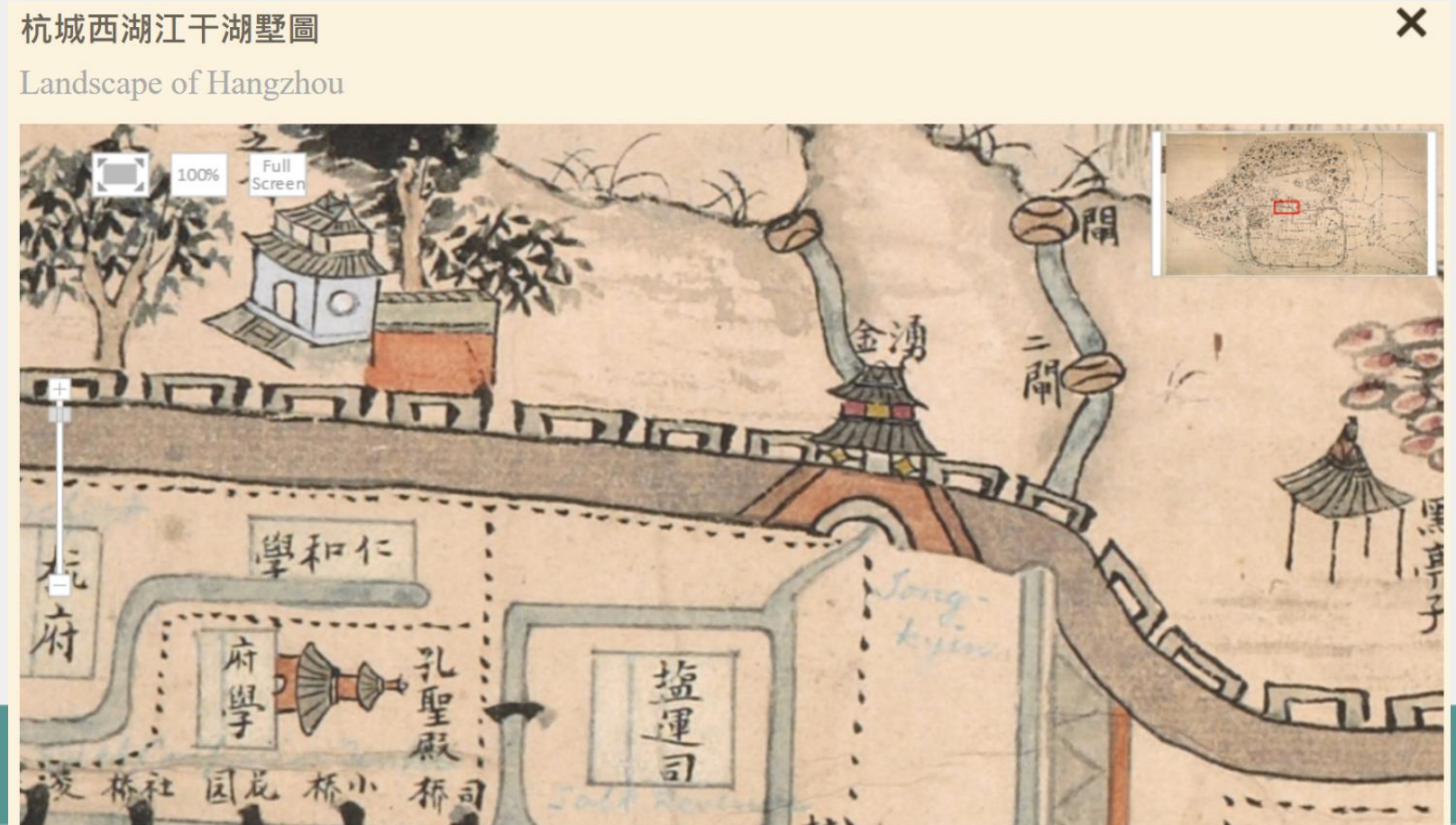
ascdc.sinica.edu.tw

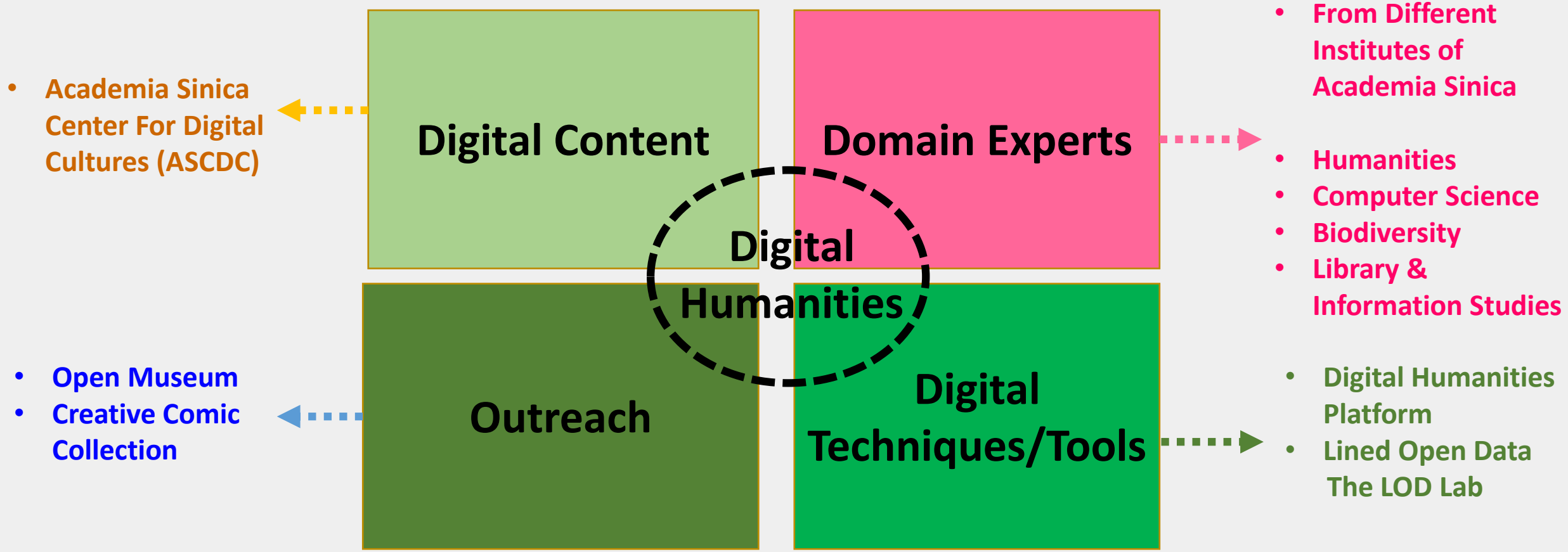
Reading Digital Atlas

http://digitalatlas.asdc.sinica.edu.tw/

keyword → maps →

full size → download



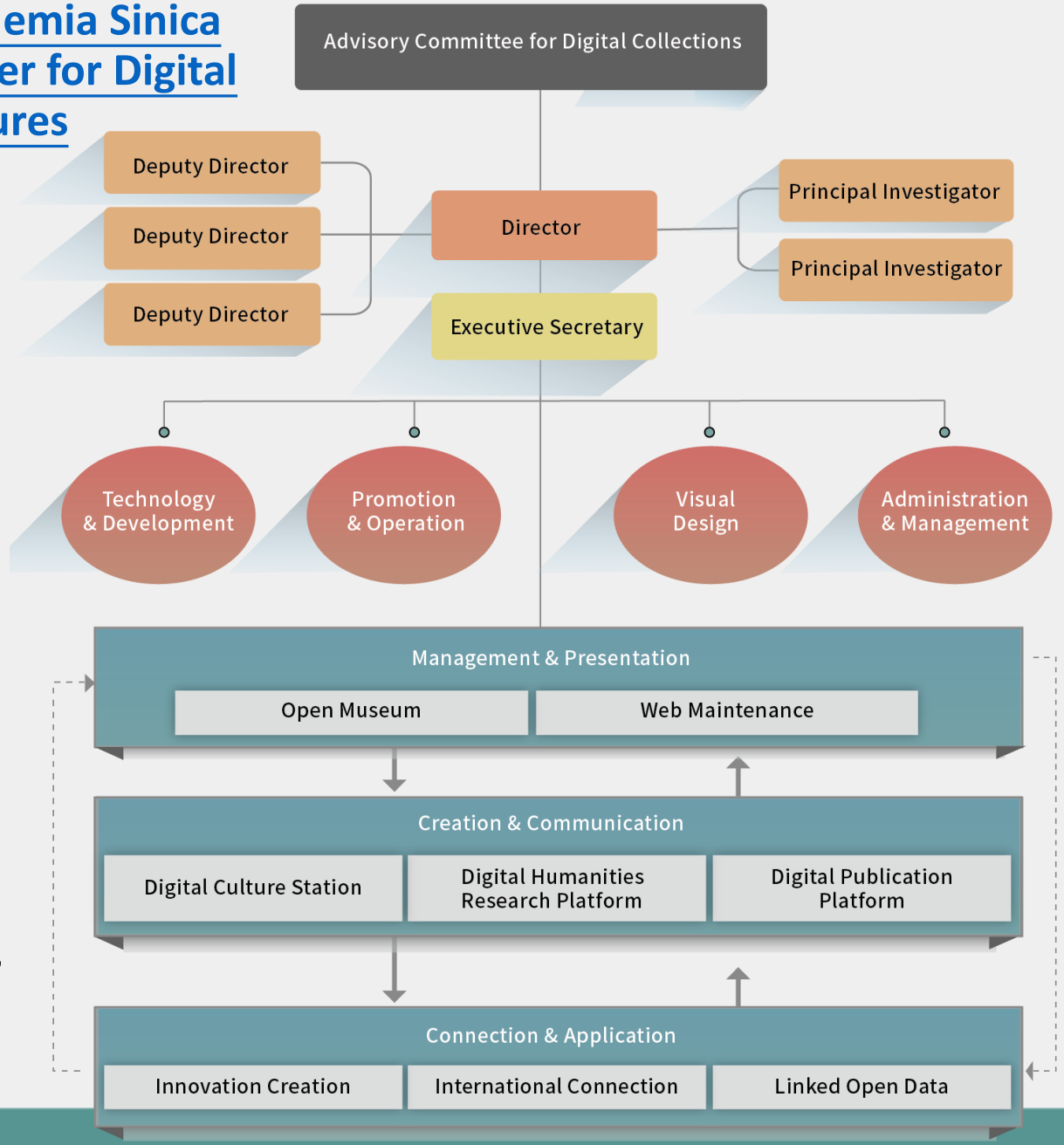

Digital Content
Domain Experts
Digital Humanities
Outreach
Digital Techniques/Tools
• Academia Sinica Center For Digital Cultures (ASCDC)
• From Different Institutes of Academia Sinica
• Humanities
• Computer Science
• Biodiversity
• Library & Information Studies
• Open Museum
• Creative Comic Collection
• Digital Humanities Platform
• Lined Open Data The LOD Lab

Organization of ASCDC

Academia Sinica Center for Digital Cultures

The Center is headed by a Director, 3 Deputy Directors and also an Executive Secretary, taken up by research fellows under appointment of the Academia Sinica President.

- Cyber-Infrastructure(The Digital Humanities Platform, The Open Museum System)
- Core Technologies (data analytics, LOD, IIIF, GIS, SNA, OCR...etc.)
- Call for Proposals (in terms of Digitization, Advanced Technologies, Collaboration)



Recent Agenda- Core Projects in ASCDC

1. Digital Humanities Research Platform
2. Linked Open Data
3. Open Museum
4. Innovative Creation (CCC project)

Academia Sinica Center for Digital Cultures

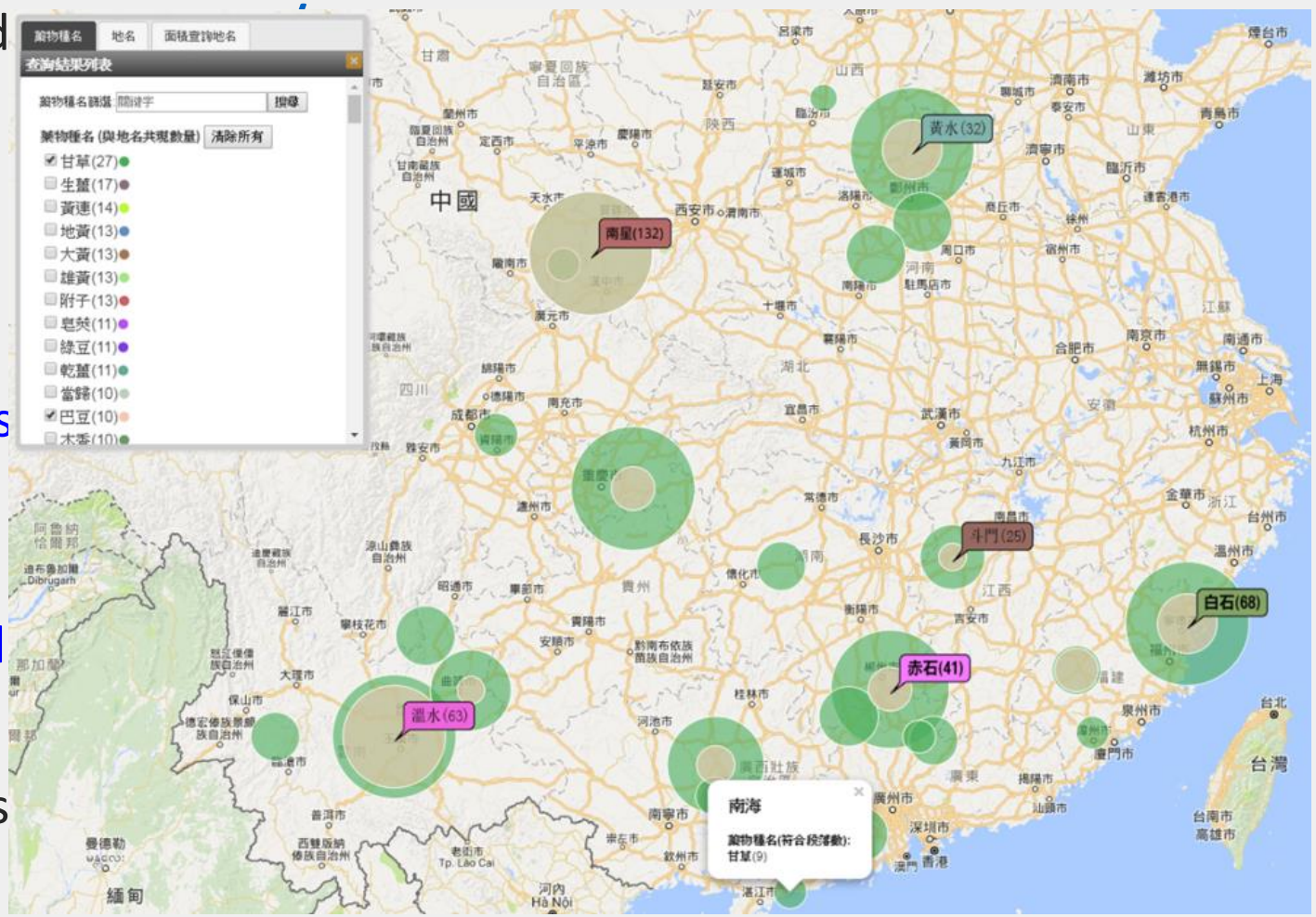
Digital Humanities Research Platform

Digital tools for scholars' research tasks

Open Platform for Digital Humanities

- Open, cloud-based text repository to enhance the research of DH; Developed as platform for online service based on the needs of scholars
- Open to individuals and institutions for co-creation and knowledge sharing
- Scholars are encouraged to upload texts to the platform and work on them together
- Platform equipped with different digital tools for text and visual analytics
- Platform will be released for all scholars in Oct. 2018

http://dh.ascdc.sinica.edu.tw/member



Characteristics in Digital Humanities Research Platform

- Integration platform of digital tools for DH research
- Search, analysis and research for (Chinese Ancient) full texts
- Personal service, eg. uploading the text and authority files
- Integrate texts from the Chinese rare books, 220 million words in current collection
- Applying the technology of natural language processing (which is collaborated with the Institute of Information Science, Academia Sinica)

1. Uploads with textual datasets
2. Rapid textual comparison
3. Analysis and statistics of authority files

Tools for DH researches

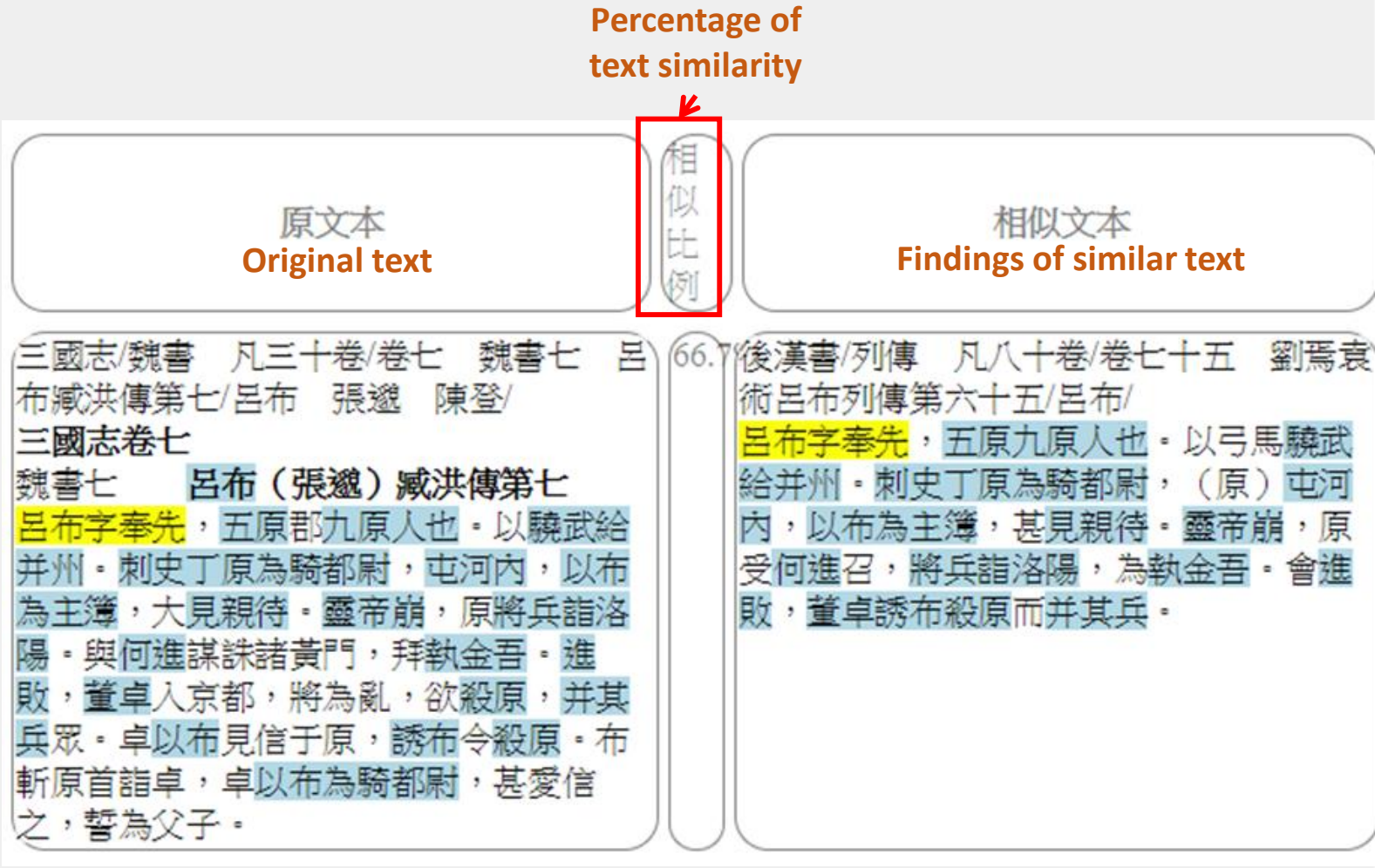
	Tools	Function
1	Text and authority files uploading and sharing	Data management
2	Text annotation and management	Data management
3	Query for regular expression	Data mining/ Research
4	Text similarity comparison	Data mining/ Research
5	Term frequency analysis (N-gram)	Data analytics
6	Association analysis	Data analytics
7	Visualization charts: word clouds, column charts, Social Network Analysis	Data visualization
8	Spatial-temporal presentation / GIS map	Data visualization

Research Tools embedded in Platform

- **Tool for text research**
  - Text annotation and management
- Annotated words can be classified in different categories

Text Similarity Comparison (Tools for data mining/research)

- Similar or identic words or passages in the text will be by automatically searched in the platform's database

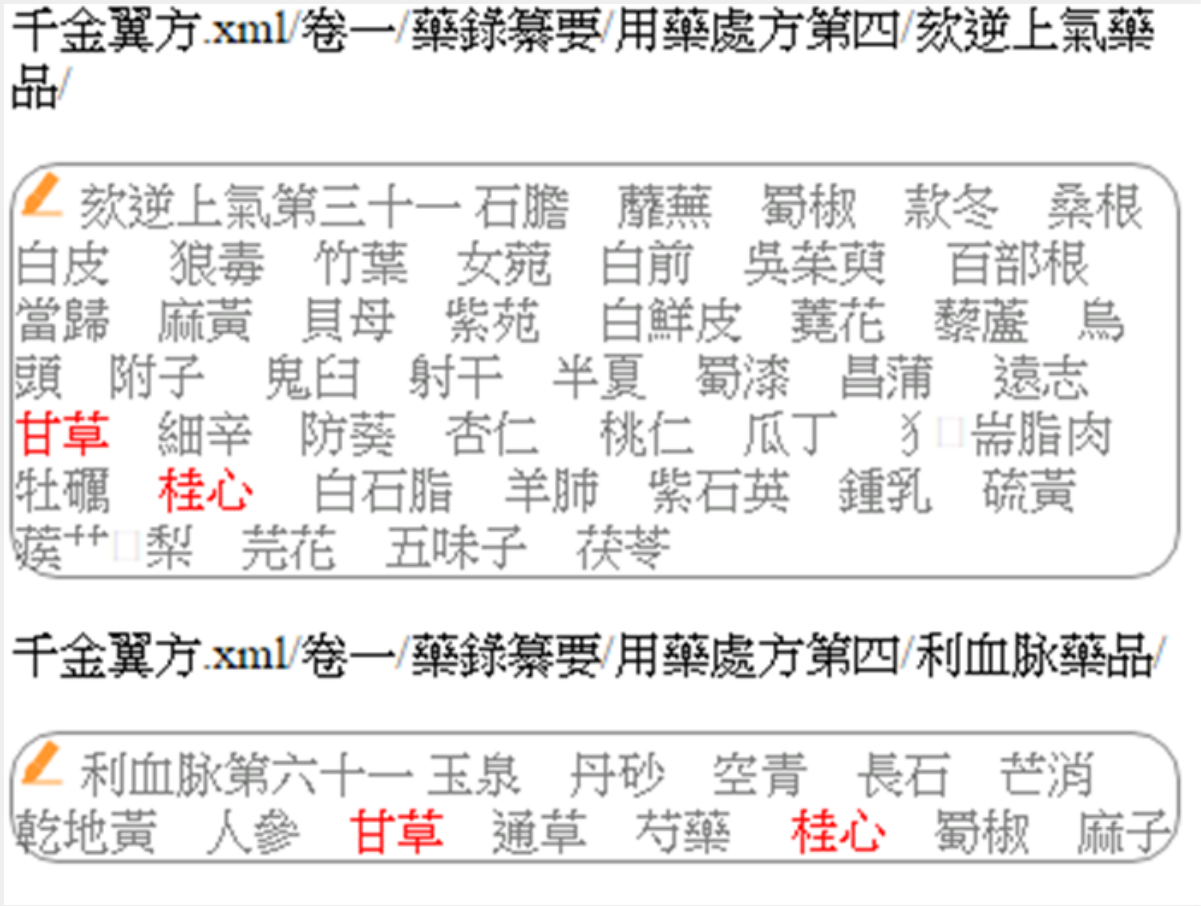


Term frequency analysis (N-gram) (Tools for Data Analysis)

- Finding the frequency of terms' occurrence in text
- Other relevant terms and terms combinations (co-occurrence analysis) can be indicated in the analysis

1. Term relation
2. Frequency
3. Terms

① 相關字	② 共現次數	③ 查詢
桂心	218	甘草 桂心
人參	217	甘草 人參
芍藥	202	甘草 芍藥
當歸	184	甘草 當歸
茯苓	161	甘草 茯苓
乾薑	151	甘草 乾薑



千金翼方.xml/卷一/藥錄纂要/用藥處方第四/欬逆上氣藥品/

欬逆上氣第三十一 石膽 蘼蕪 蜀椒 款冬 桑根白皮 狼毒 竹葉 女菀 白前 吳茱萸 百部根當歸 麻黃 貝母 紫菀 白鮮皮 蕘花 藜蘆 烏頭 附子 鬼臼 射干 半夏 蜀漆 昌蒲 遠志甘草 細辛 防葵 杏仁 桃仁 瓜丁 □□崙脂肉牡礪 桂心 白石脂 羊肺 紫石英 鍾乳 硫黃□□梨 芫花 五味子 茯苓

千金翼方.xml/卷一/藥錄纂要/用藥處方第四/利血脉藥品/

利血脉第六十一 玉泉 丹砂 空青 長石 芒消乾地黃 人參 甘草 通草 芍藥 桂心 蜀椒 麻子

Term's occurrence of "Liquorice" (甘草) in the Chinese medical books and its co-relation with other terms of Chinese medical plants

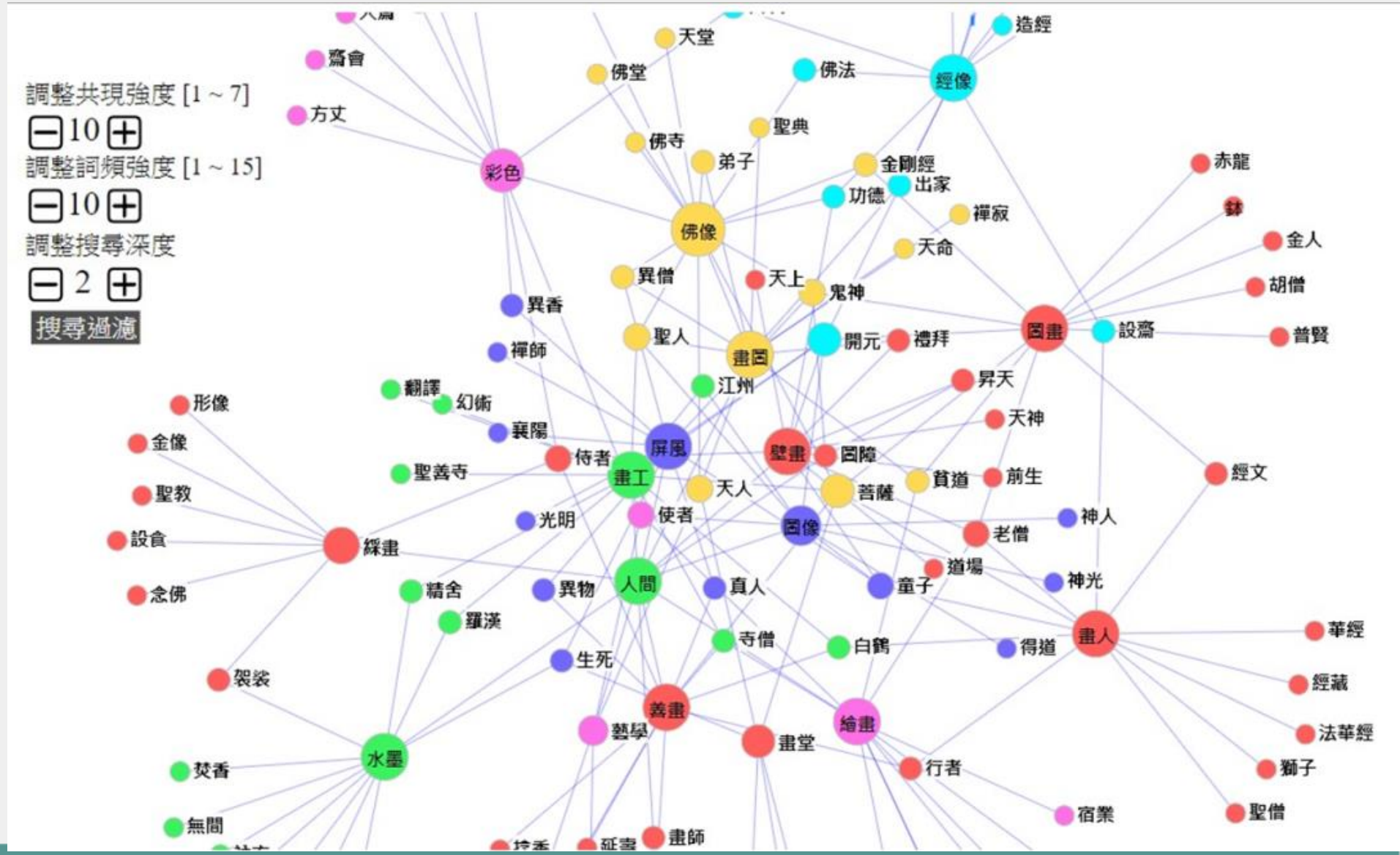
Tool for data visualization

關鍵字	平均頻率	平均值
道士	769	100
太平	599	99
人間	367	99
弟子	363	99
使者	345	99
神仙	292	99
老人	284	99
不食	186	99
童子	161	99
仙人	154	99
神人	147	99
鬼神	147	99
王母	146	99
功德	140	98
紫衣	137	98

Word Cloud

Visualization tool for overviewing the term frequency of co-occurrence

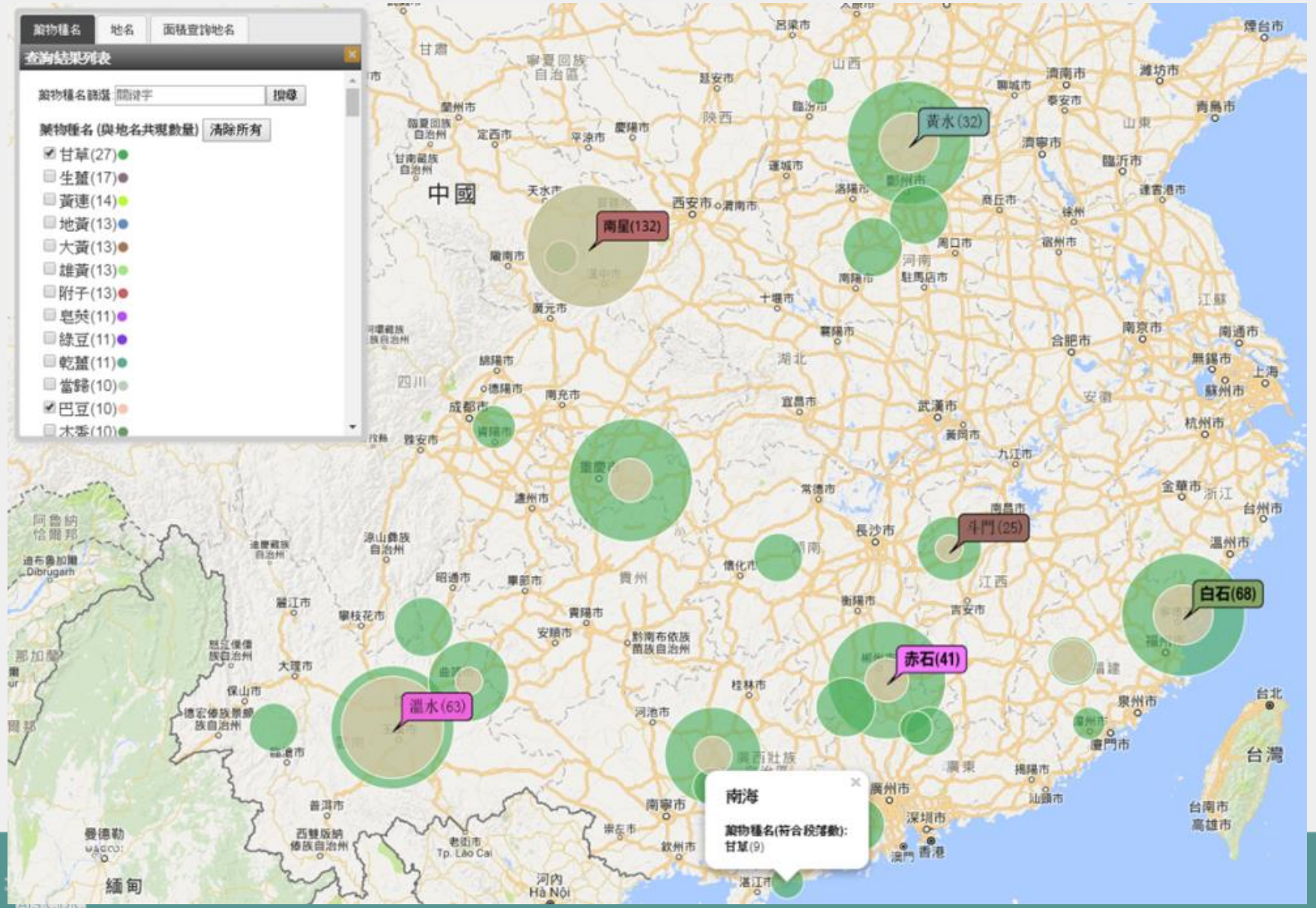
Tool for data visualization



Network figure of Terms' co-relations

Co-relation between the terms heading under subjects of "paintings" and "Buddhism"

Tool for data visualization



Spatial presentation in GIS map

Plants' growing location in map

Selected Research Projects with DH Research Platform

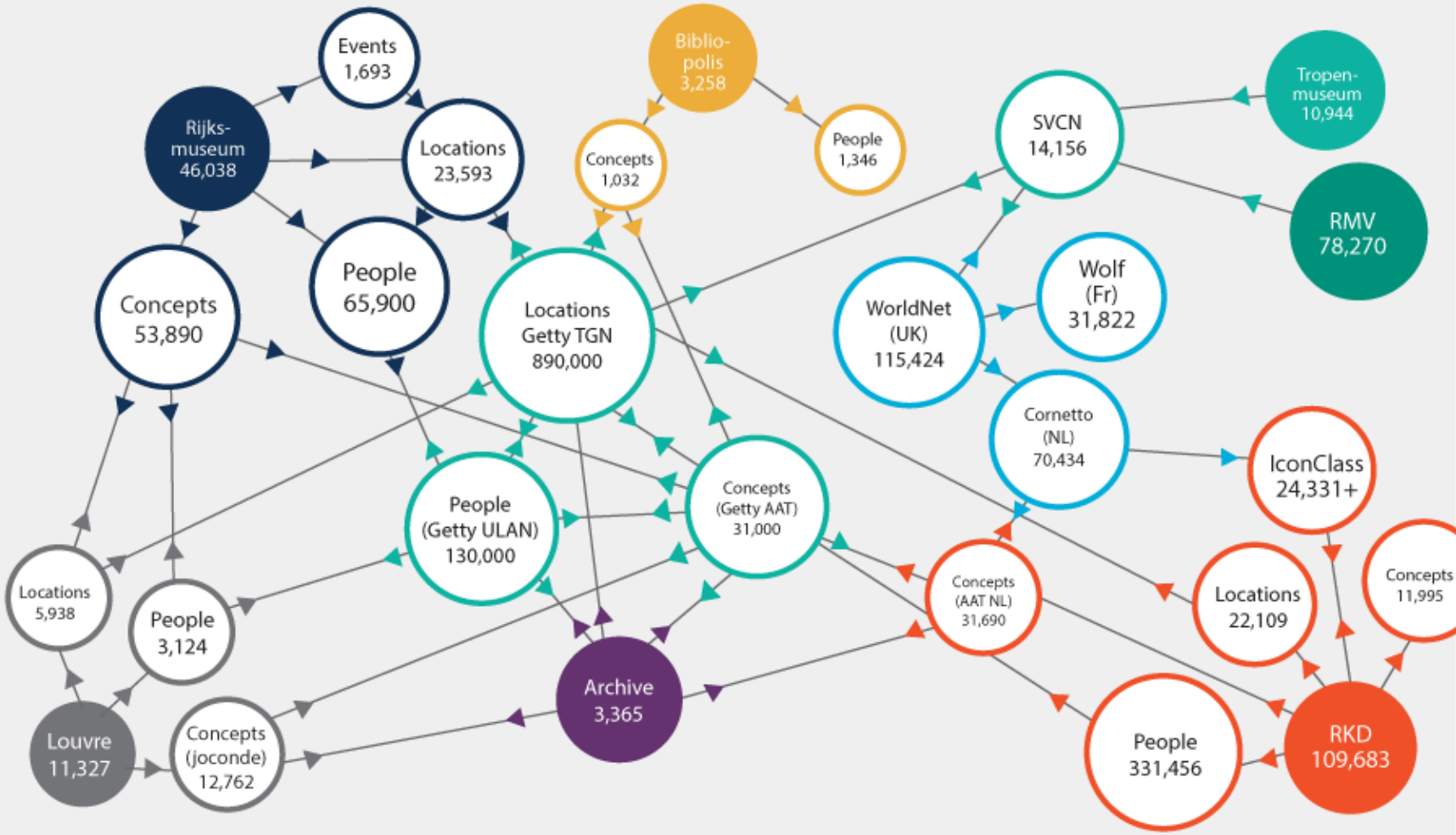
	Project Title	Institute	Topics
1	Gender, Media and War: Guomindang woman's Social networks during the war of resistance against Japan 性別、媒體與戰爭：抗戰時期國民黨婦女的社會網絡	Institute of Modern History, Academia Sinica	Gender study
2	Preliminary study of local elite networks in Japanese colonial Taiwan: the case of IHT Diary Knowledge Bank日治時期臺灣菁英的社會網絡初探：以楊肇嘉（1892-1976）與中研院臺史所「日記知識庫」為例	Institute of Taiwan History, Academia Sinica	Cultural history
3	Digital editions for the study of early medieval China 中古文學研究的數位版本	Institute of Chinese Literature and Philosophy, Academia Sinica	Literature
5	Pains and fears: DH studies on the descriptions of illness made by the Chinese Buddhist monks 病苦與死懼：歷代僧傳疾病敘述的數位人文研究	Institute of Chinese Literature and Philosophy, Academia Sinica	Religion history

http://lodlab.ascdc.tw/en/

LODLab

Academia Sinica Center for Digital Cultures

The laboratory works on research and practical contributions toward linked open data. We are currently **concentrating on Taiwan's digital** collections, with steady progress being made on distribution and conversion of metadata to structured semantic datasets. This allows meaningful linkages to be established between data worldwide, supplying researchers and the public with a more complete and multi-perspective context for data, deepening the bases of exploration, analysis, and research.



LOD as method to preserve digital cultural assets

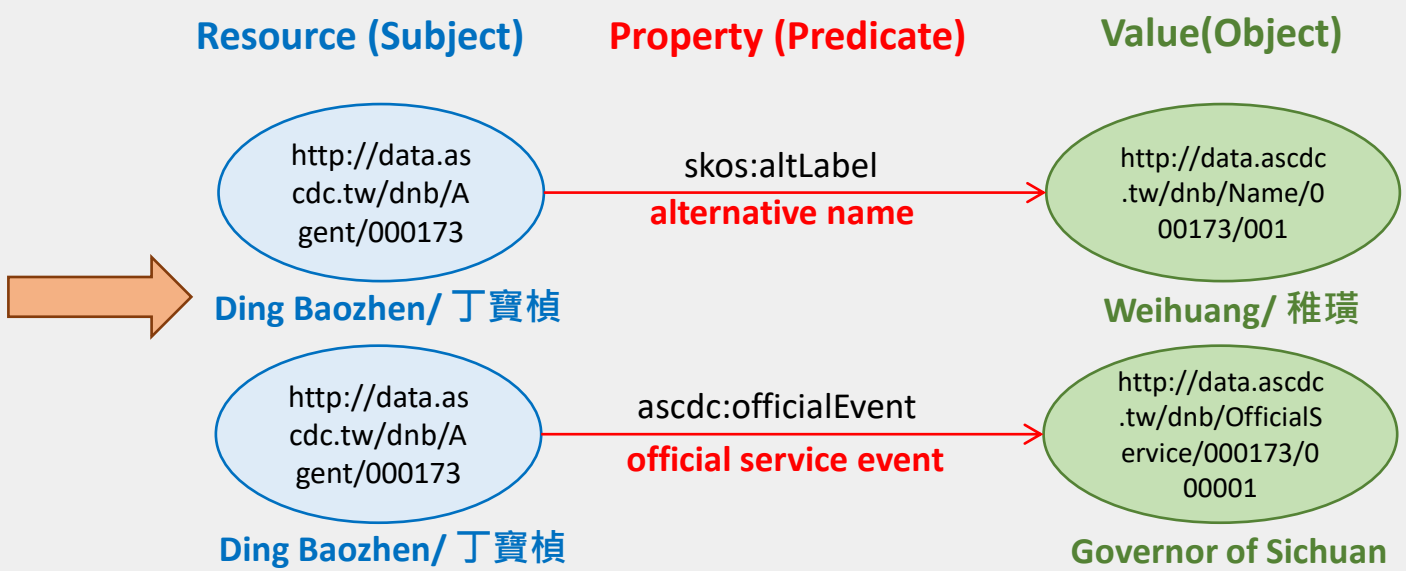
- Core works:

Transformation of the traditional metadata into semantic data model, composed of sets of triples endowed with URIs: Subject – Predicate – Object

- Efficient way for the permanent preservation of digital cultural assets in the world of Web 3.0

	Field	Field Vale
1	Name	Ding Baozhen/ 丁寶楨
2	Alternative name	Weihuang/ 稚璜
3	Identifier	000173
4	Biographic description	Born in Pingyuan, Guizhou Province in 1820, Ding was appointed a government official in 1854 …
5	Event/ official service	Governor of Sichuan Province

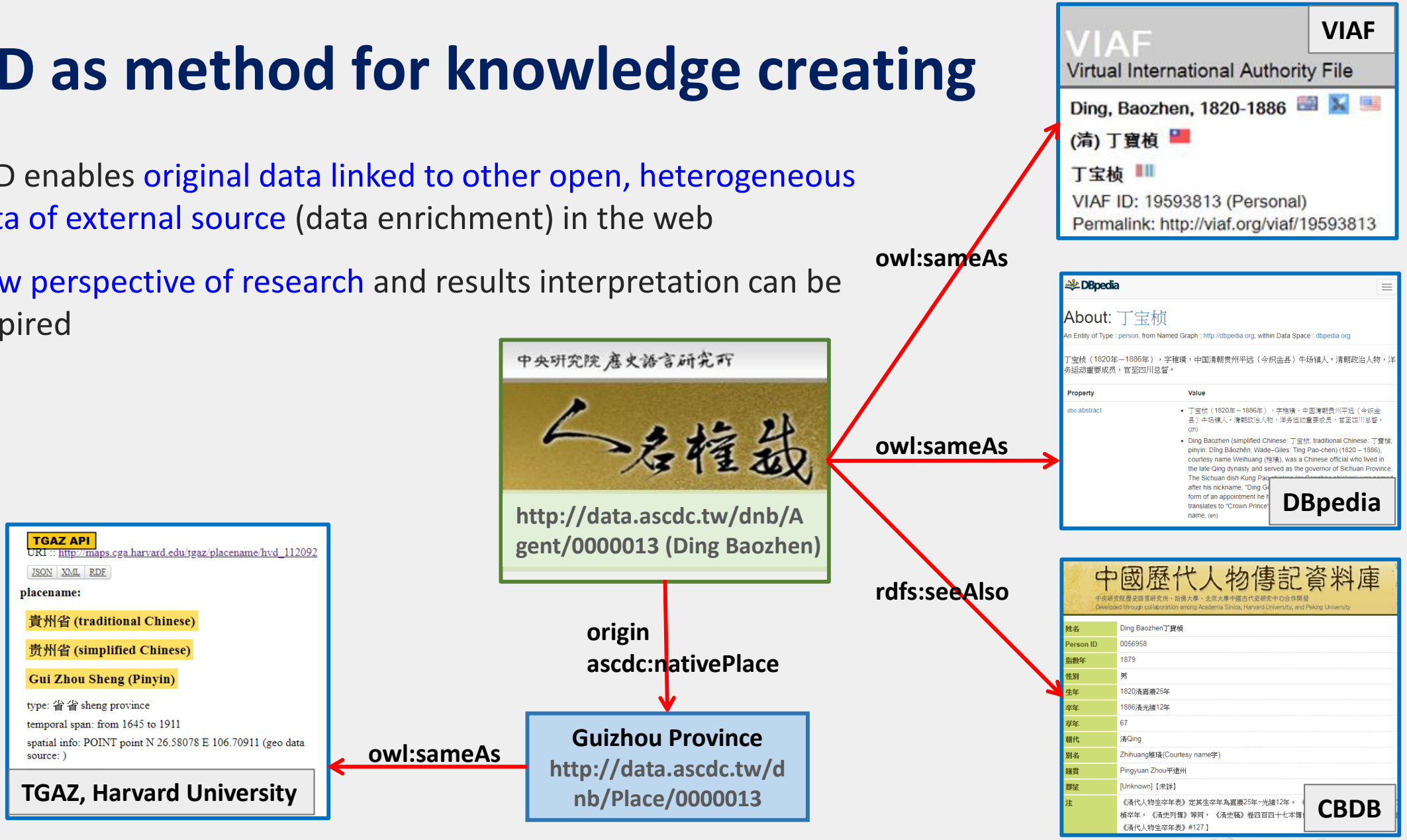
Metadata table



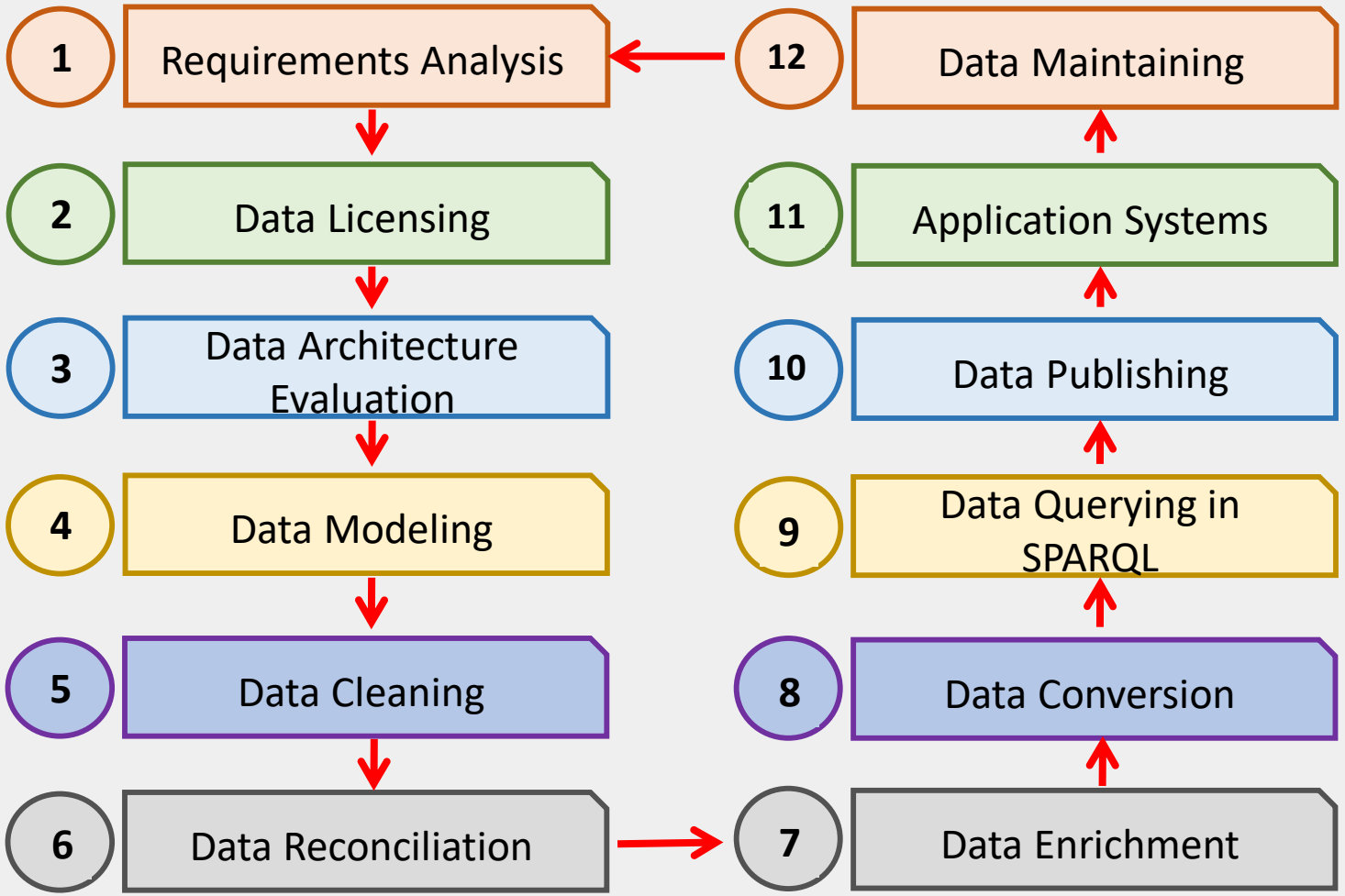
Semantic data model

LOD as method for knowledge creating

- LOD enables original data linked to other open, heterogeneous data of external source (data enrichment) in the web
- New perspective of research and results interpretation can be inspired



The LOD Lifecycle Model



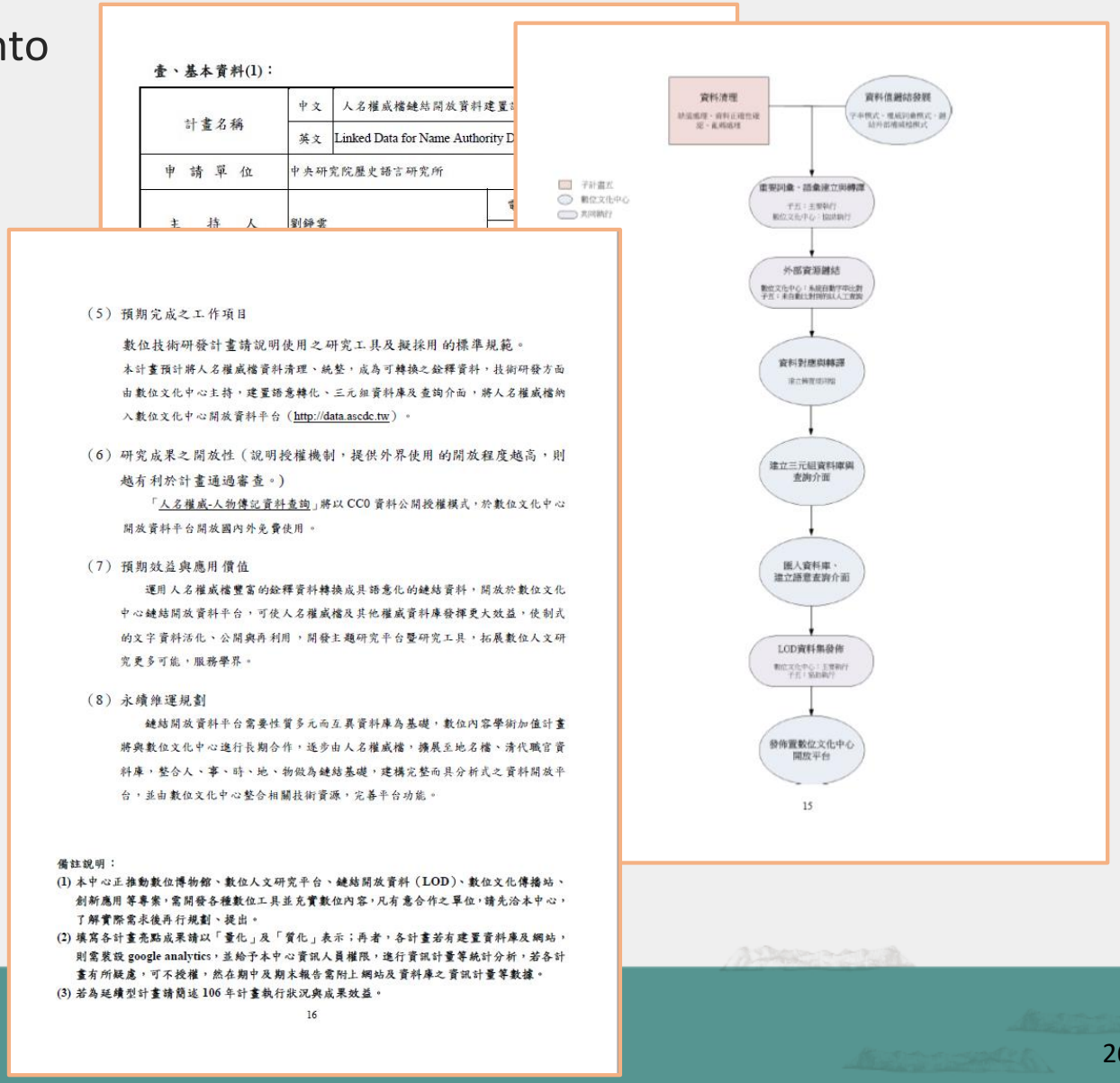
LOD Lifecycle- (1)Requirements Analysis

- To ensure the purposes of metadata conversion into LOD and requirements to the future application

- From Web of Document to Web of Data

- As a Reference Tool v.s. As a Research Tool

- Understanding the requirements by reading and analyzing the project proposal
- Case: LOD project-Database of Names and Biographies (DNB)



壹、基本資料(1)：

計畫名稱	中文	人名權威檔鏈結開放資料建置
	英文	Linked Data for Name Authority D
申請單位	中央研究院歷史語言研究所	
主持人	劉錚雲	

（5）預期完成之工作項目

數位技術研發計畫請說明使用之研究工具及擬採用的標準規範。

本計畫預計將人名權威檔資料清理、統整，成為可轉換之詮釋資料，技術研發方面由數位文化中心主持，建置語意轉化、三元組資料庫及查詢介面，將人名權威檔納入數位文化中心開放資料平台（http://data.ascdc.tw）。

（6）研究成果之開放性（說明授權機制，提供外界使用的開放程度越高，則越有利於計畫通過審查。）

「人名權威-人物傳記資料查詢」將以 CC0 資料公開授權模式，於數位文化中心開放資料平台開放國內外免費使用。

（7）預期效益與應用價值

運用人名權威檔豐富的詮釋資料轉換成具語意化的鏈結資料，開放於數位文化中心鏈結開放資料平台，可使人名權威檔及其他權威資料庫發揮更大效益，使制式的文字資料活化、公開與再利用，開發主題研究平台暨研究工具，拓展數位人文研究更多可能，服務學界。

（8）永續維運規劃

鏈結開放資料平台需要性質多元而互異資料庫為基礎，數位內容學術加值計畫將與數位文化中心進行長期合作，逐步由人名權威檔，擴展至地名檔、清代職官資料庫，整合人、事、時、地、物做為鏈結基礎，建構完整而具分析式之資料開放平台，並由數位文化中心整合相關技術資源，完善平台功能。

備註說明：

(1) 本中心正推動數位博物館、數位人文研究平台、鏈結開放資料（LOD）、數位文化傳播站、創新應用等專案，需開發各種數位工具並充實數位內容，凡有意合作之單位，請先洽本中心，了解實際需求後再行規劃、提出。

(2) 填寫各計畫亮點成果請以「量化」及「質化」表示；再者，各計畫若有建置資料庫及網站，則需裝設 google analytics，並給予本中心資訊人員權限，進行資訊計量等統計分析，若各計畫有所疑慮，可不授權，然在期中及期末報告需附上網站及資料庫之資訊計量等數據。

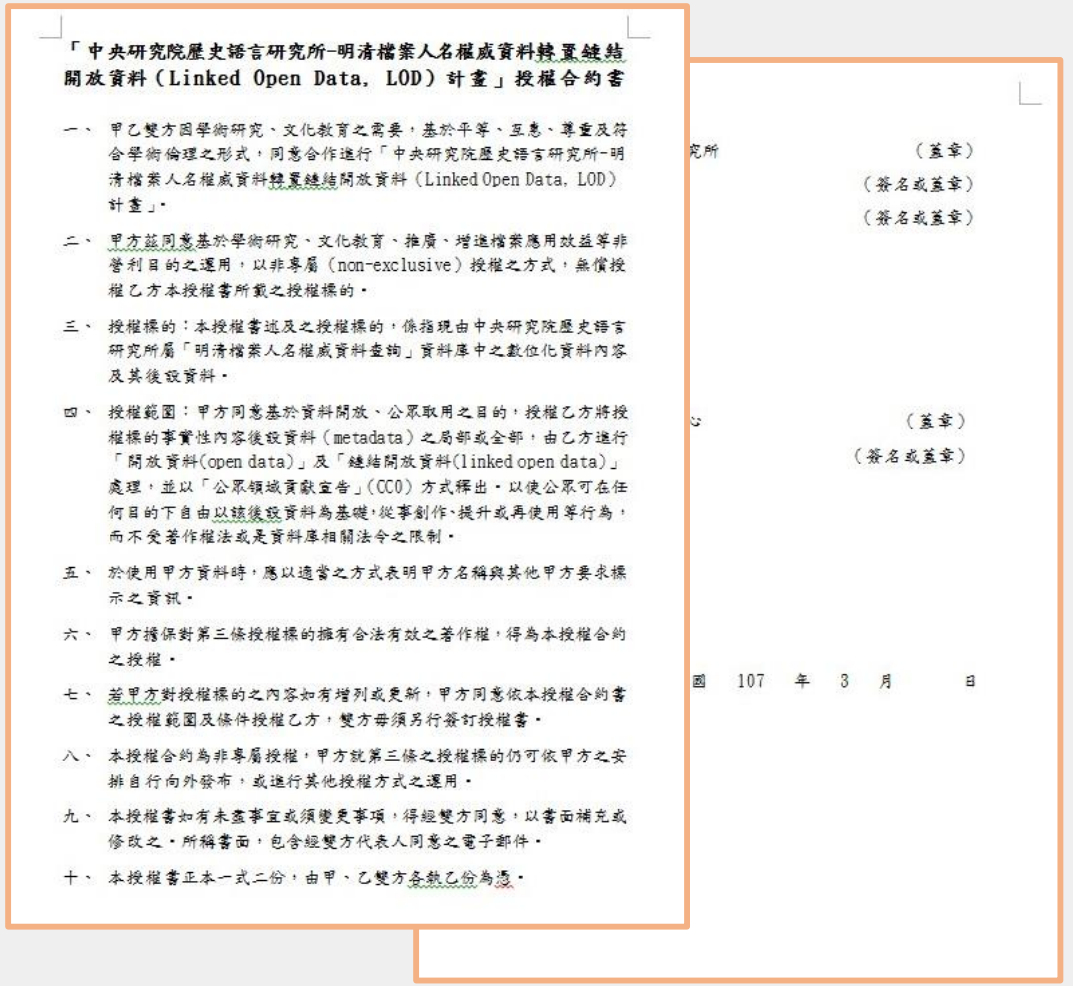
(3) 若為延續型計畫請簡述 106 年計畫執行狀況與成果效益。

16

15

LOD Lifecycle- (2)Data Licensing

- To ensure and acquire the license permission for data re-consumption



「中央研究院歷史語言研究所-明清檔案人名權威資料轉置鏈結開放資料（Linked Open Data, LOD）計畫」授權合約書

一、 甲乙雙方因學術研究、文化教育之需要，基於平等、互惠、尊重及符合學術倫理之形式，同意合作進行「中央研究院歷史語言研究所-明清檔案人名權威資料轉置鏈結開放資料（Linked Open Data, LOD）計畫」。

二、 甲方茲同意基於學術研究、文化教育、推廣、增進檔案應用效益等非營利目的之運用，以非專屬（non-exclusive）授權之方式，無償授權乙方本授權書所載之授權標的。

三、 授權標的：本授權書述及之授權標的，係指現由中央研究院歷史語言研究所屬「明清檔案人名權威資料查詢」資料庫中之數位化資料內容及其後設資料。

四、 授權範圍：甲方同意基於資料開放、公眾取用之目的，授權乙方將授權標的事實性內容後設資料（metadata）之局部或全部，由乙方進行「開放資料(open data)」及「鏈結開放資料(linked open data)」處理，並以「公眾領域貢獻宣告」(CC0) 方式釋出，以使公眾可在任何目的下自由以該後設資料為基礎，從事創作、提升或再使用等行為，而不受著作權法或是資料庫相關法令之限制。

五、 於使用甲方資料時，應以適當之方式表明甲方名稱與其他甲方要求標示之資訊。

六、 甲方擔保對第三條授權標的擁有合法有效之著作權，得為本授權合約之授權。

七、 若甲方對授權標的之內容如有增列或更新，甲方同意依本授權合約書之授權範圍及條件授權乙方，雙方毋須另行簽訂授權書。

八、 本授權合約為非專屬授權，甲方就第三條之授權標的仍可依甲方之安排自行向外發布，或進行其他授權方式之運用。

九、 本授權書如有未盡事宜或須變更事項，得經雙方同意，以書面補充或修改之。所稱書面，包含經雙方代表人同意之電子郵件。

十、 本授權書正本一式二份，由甲、乙雙方各執乙份為憑。

（蓋章）
（簽名或蓋章）
（簽名或蓋章）
（蓋章）
（簽名或蓋章）
107 年 3 月 日

License permission of DNB data conversion

Purpose of data licensing

1. The descriptive metadata in the LOD dataset should be released under CC-BY or CC0 (in optimal situation)
2. Through announcing the data license, the data can be further freely reused, re-created by the public. A better distribution of the study results can be thus enhanced.

LOD Lifecycle- (3)Data Architecture Evaluation

- To analyze the data structure, data content and data value of the original metadata. In the meanwhile, to review the currently reusable ontologies and authority files, which can be applied as basis for data modeling

To analyze the quality of data content, whether the row data is well structured

To check whether the data in a single field content too many varied, heterogeneous information

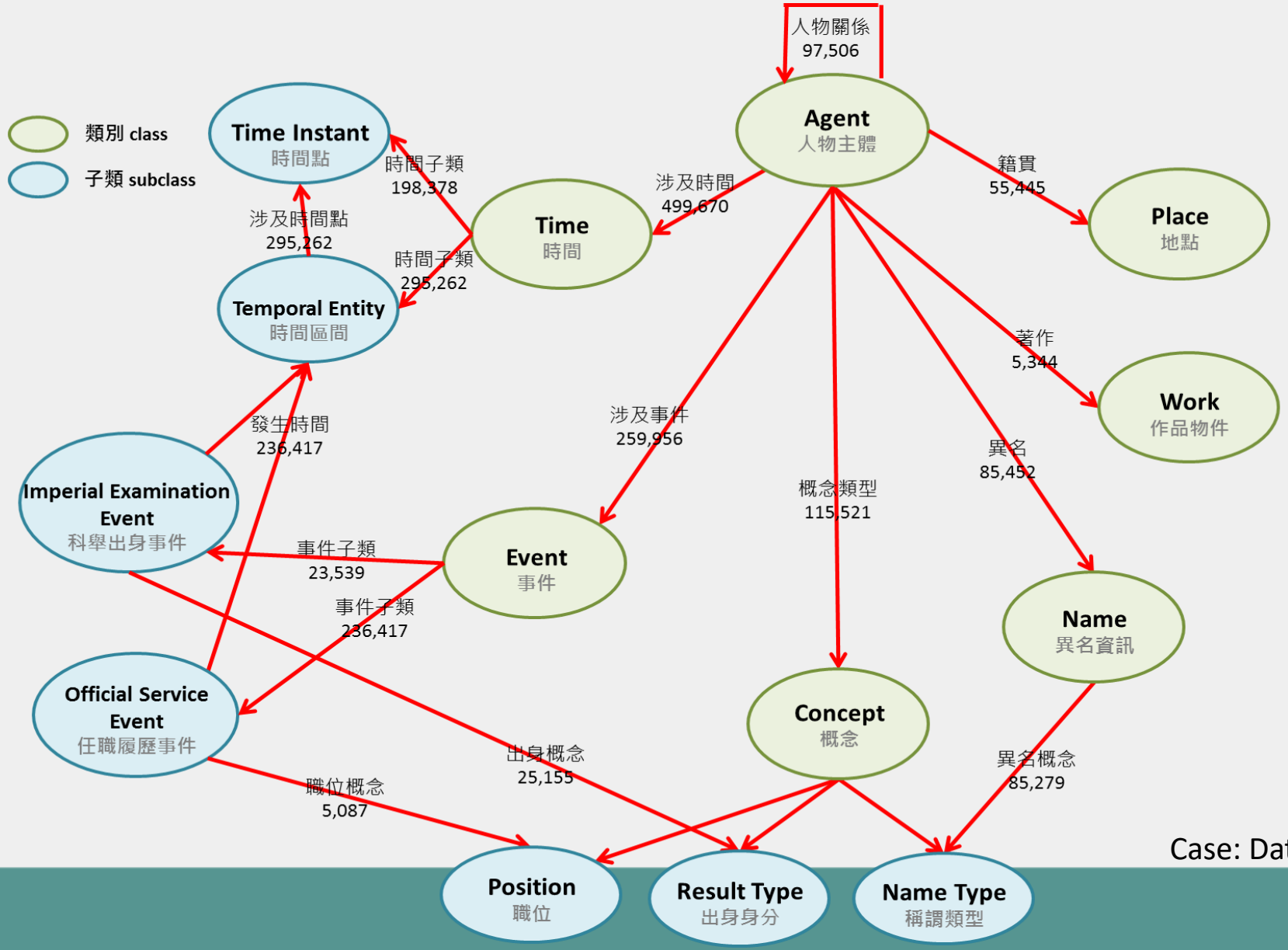
	A	B	C	D	E	F	G	H	I	J	K	L	M	
2	系統號	姓名	朝代	中曆生卒年	西曆生卒年	籍貫(原地名)	籍貫(新地名)	八旗(說明/隸屬)	經度	緯度	異名(種類/名稱/出處)AA	關連		傳略
3	NO000000001	(富察)傅鼐	清	康熙[16]年-乾隆3年	[1677]-1738			^/滿洲鑲白旗			其他^富鼐^欽定八旗通志322卷1; 字^	富察額色泰(祖)；富察喝爾漢(父)；富察昌		傅鼐,字
4	NO000000002	邵洪	清	乾隆8年-嘉慶16年	1743-1811	浙江省-寧波府-鄞縣	浙江省寧波市		121.5	29.8	字^海度^國立故宮博物院圖書文獻處	邵基(祖父)；邵鐸(父)；邵桓(子)		重修浙
5	NO000000003	刁承祖	清	康熙11年-乾隆4年	1672-1739	直隸省-保定府-祁州	河北省安國縣		115.3	38.4	字^步武^碑傳集5冊82卷586; 漢語拼音	刁克俊(曾祖父)；刁包(祖父)；刁再濂(父)；		沈大成
6	NO000000004	邵基	清	康熙26年-乾隆2年	1687-1737	浙江省-寧波市-鄞縣	浙江省寧波市		121.5	29.8	字^學址^國立故宮博物院圖書文獻處	邵淡(父)；邵鐸(子)；邵洪(孫)；邵桓(曾孫)		重修浙
7	NO000000005	于成龍	明－清	萬曆45年-康熙23年	1617-1684	山西縣-汾州府-永寧州	山西省呂梁市離石區		111.150733	37.517641	字^北溟^國立故宮博物院圖書文獻處	于時煌(父)；于廷翼(子)；于廷勳(子)；于廷		于成龍,
8	NO000000006	(瓜勒佳)三德	清	？-乾隆53年	？-1788			^/滿洲鑲紅旗			證^昭毅^清代職官年表4冊3130; 羅馬拼音^sande^;			楊于果
9	NO000000007	金德瑛	清	康熙40年-乾隆27年	1701-1762	浙江省-杭州府-仁和縣	浙江省杭州市		120.15507	30.274085	字^汝白^國立故宮博物院圖書文獻處	金敦順(祖父)；金聖緒(父)；金德鎔(兄)；		金德瑛,
10	NO000000008	(章佳)阿桂	清	康熙56年-嘉慶2年	1717-1797			原滿洲正藍旗^/滿洲正白旗			字^廣廷^清人室名別稱字號索引下冊	章佳雅爾泰(曾祖父)；章佳阿恩哈(祖父)；		阿桂,字
11	NO000000009	姚文然	明－清	泰昌1年-康熙17年	1620-1678	南京-安慶府-桐城縣	安徽省安慶市桐城市		116.97412	31.0358	字^弱侯^國立故宮博物院圖書文獻處	姚文熊(弟)；姚士基(子)；姚士塾(子)；姚		姚文然,
12	NO000000010	于成龍	明－清	崇禎11年-康熙39年	1638-1700	[遼東都指揮使司-蓋州衛](碑	[遼寧省營口市蓋縣]	^/[漢軍鑲黃旗][漢軍	122.4	40.4	字^振甲^國立故宮博物院圖書文獻處	于永禎(子)；于永裕(子)；于永世(子)；于		王士禎
13	NO000000011	(章佳)尹繼善	清	康熙[33]年-乾隆36年	[1694]-1771			^/滿洲鑲黃旗			字^元長^國立故宮博物院圖書文獻處	章佳尹泰(父)；章佳尹立善(兄)；章佳慶桂		尹繼善,
14	NO000000012	(愛新覺羅)允祥	清	康熙25年-雍正8年	1686-1730			^/滿洲			其他^胤祥^國立故宮博物院圖書文獻	愛新覺羅皇太極(曾祖父)；愛新覺羅福臨(		鄧邦述
15	NO000000013	尹會一	清	康熙30年-乾隆13年	1691-1748	直隸省-保定府-博野縣	河北省保定市博野縣		115.464402	38.457432	字^元孚^國立故宮博物院圖書文獻處	尹先知(曾祖父)；尹澤升(祖父)；[尹元弼][		尹會一,
16	NO000000014	施琅	明－清	天啟1年-康熙35年	1621-1696	福建承宣布政使司-泉州府-晉	福建省泉州市	^/改隸漢軍鑲黃旗	118.6	24.9	字^琢公^國立故宮博物院圖書文獻處	施大宣(父)；施武毅(叔父)；施顯(弟)；施		連橫施
17	NO000000015	胡林翼	清	嘉慶17年-咸豐11年	1812-1861	湖南省-長沙府-益陽縣	湖南省益陽市		112.35518	28.55386	字^貺生^國立故宮博物院圖書文獻處	胡漢清(始祖)；胡民典(高祖)；胡多吉(曾祖		胡林翼,
18	NO000000016	孔有德	明－清	萬曆30年-順治9年	1602-1652	遼東都指揮使司	遼寧省遼陽市	^/漢軍正紅旗	123.2	41.3	證^武壯^國立故宮博物院圖書文獻處	孔文龍(養父)；孔廷訓(子)；孔四貞(女兒)		八旗通
19	NO000000017	胡寶瑔	清	康熙33年-乾隆28年	1694-1763	江蘇省-松江府-青浦縣	上海市青浦區		121.124178	31.150681	其他^胡金蘭^廣清碑傳集7卷483; 字^	胡希烈(祖)；胡廷對(父)；胡嘉會(兄)；胡		王昶胡
20	NO000000018	凌如煥	清	康熙20年-乾隆[13][1	1681-[1748][1749]	江蘇省-松江府-上海縣	上海市黃浦區老城廂		121.4683	31.2287	字^榆山^廣清碑傳集7卷438; [不詳]^	凌應蘭(子)；凌應豹(子)		袁枚兵
21	NO000000019	土國寶	明－清	？-順治8年	？-1651	山西承宣布政使司-大同府-[	山西省大同市		113.300126	40.076816	漢語拼音^Tu Guobao^			土國寶
22	NO000000020	孫士毅	清	康熙59年-嘉慶1年	1720-1796	浙江省-杭州府-仁和縣	浙江省杭州市		120.15507	30.274085	字^智治^國立故宮博物院圖書文獻處	孫玉亭(父)；孫士達(兄)；孫士倫(兄)；孫		袁枚太
23	NO000000021	方苞	清	康熙7年-乾隆14年	1668-1749	安徽省-安慶府-桐城縣	安徽省安慶市桐城市		116.97412	31.0358	字^靈皋^國立故宮博物院圖書文獻處	方象乾(曾祖)；方熾(祖)；方仲舒(父)；方		方苞,字
24	NO000000022	方觀承	清	康熙37年-乾隆33年	1698-1768	安徽省-安慶府-桐城縣	安徽省安慶市桐城市		116.97412	31.0358	字^遐穀^國立故宮博物院圖書文獻處	方登嶧(祖父)；方式濟(父)；方觀永(兄)；		方觀承,
25	NO000000023	孫嘉淦	清	康熙22年-乾隆18年	1683-1753	山西省-太原府-興縣	山西省呂梁市興縣		111.12774	38.462356	字^錫公^國立故宮博物院圖書文獻處	孫守意(曾祖父)；孫世蕙(祖父)；孫天繡(父		孫嘉淦,
26	NO000000024	王士禎	明－清	崇禎7年-康熙50年	1634-1711	山東承宣布政使司-濟南府-新	山東省桓臺縣新城鎮		117.951772	36.956046	改名^王士正^國立故宮博物院圖書文	王貴(遠祖)；王之坦(曾祖父)；王象晉(祖父		欽定國
27	NO000000025	孫蘭	明－清	[不詳]	[不詳]	[不詳]	[不詳]				漢語拼音^Sun Lan^			

Case: original metadata of DNB

To analyze whether the data values can be linked to the external authority resources (eg. TGN/ Geonames)

LOD Lifecycle- (4)Data Modeling

design the ontology/
semantic data model



Case: Data model of DNB

LOD Lifecycle- (5)Data Cleaning

- To transform the original data to structured one and correct the error data for optimization of the data quality

1	35253筆	http://archive.ihp.sinica.edu.tw/ttsweb/html_name/search.php						
2	**S**	**欄位類型**	**P**	**轉置規則說明/ 補充說明**	**案例**	**O**	**T**	**L**
3	http://data.ascdc.tw/dnb/Agent/系統號	系統號	dcterms:identifier	文字直接轉出	NO000000105		Literal	en
4	http://data.ascdc.tw/dnb/Agent/系統號	姓名	skos:prefLabel	文字直接轉出	翁方綱		Literal	zh
5	http://data.ascdc.tw/dnb/Agent/系統號	朝代	ascdc:dynasty	第一階段: 先以原字串轉出。如欄位值為"？"、"[不詳]"者先不轉出。 第二階段: 1. 以"–"進行斷詞，斷詞後的詞彙與AAT或DBpedia對應(可參酌時間控制表_朝代)，如可對應者，鏈結至DBpedia詞彙URI，如不行文字轉出 2. 斷詞後的資料為"？"、"[不詳]"者先不轉出。	明–清	明–清 (第一階段) http://dbpedia.org/resource/Ming_dynasty (第二階段) http://dbpedia.org/resource/Qing_dynasty (第二階段)	第一階段:Literal 第二階段: URI/ Literal	zh zh
6	http://data.ascdc.tw/dnb/Agent/系統號	中曆生卒年	ascdc:atTime	先以原字串轉出(避免資料值中一些不詳或不確定之資料值套用time ontology的問題)。	？-康熙？年 [不詳]	？-康熙？年 [不詳]	Literal	zh
7	http://data.ascdc.tw/dnb/Agent/系統號	中曆生卒年(時間區段)	time:hasTime	1. 資料值為"[不詳]"或單一資料值中有兩個"?"(如乾隆？年-？)者先不轉出 2. 剩餘O值轉為class並賦予時間區段URI流水編號。TimeInterval後為流水號http://data.ascdc.tw/dnb/TimeInterval/系	雍正11年-嘉慶23年	http://data.ascdc.tw/dnb/TimeInterval/系統號/001	URI	
8	http://data.ascdc.tw/dnb/TimeInterval/000000001	中曆生卒年(生年區段、卒	time:hasBeginning	1. 將上一步驟中，去除不明確資料值後的O值，以"-"進行斷詞。 2. 斷出後，"-"前的中曆年資訊歸於生年欄位，並與時間控制表中"年"(year)的資訊對應、轉為class並賦予time ontologyURI流水編號，如http://data.ascdc.tw/time/chi/year/00001。	雍正11年	http://data.ascdc.tw/time/chi/year/00001 (雍正11年)	URI	
9	http://data.ascdc.tw/dnb/TimeInterval/000000001	中曆生卒年(生年區段、卒	time:hasEnd	1. 將上一步驟中，去除不明確資料值後的O值，以"-"進行斷詞。 2. 斷出後，"-"後的中曆年資訊歸於卒年欄位，並與時間控制表中"年"(year)的資訊對應、轉為class並賦予time ontologyURI流水編號，如http://data.ascdc.tw/time/chi/year/00002。	嘉慶23年	http://data.ascdc.tw/time/chi/year/00002 (嘉慶23年)	URI	
10	http://data.ascdc.tw/time/chi/era/001	中曆生卒年(生年區段、卒	time:inside	1. 將上一步驟中，所有建立好的生年class及卒年class與時間控制表中"年號"(era)的資訊對應 2. 利用屬性time:inside將對應到的年號class/URI與"年"的class/URI鏈結	雍正11年	http://data.ascdc.tw/time/chi/year/00001 (雍正11年)	URI	
11	http://data.ascdc.tw/time/chi/era/001	中曆生卒年(生年區段、卒	time:intervalDuring	1. 將上一步驟中對應到的"年號"(era)class/URI與時間控制表中"皇帝統治期間"(emperor)的資訊進行對應。 2. 利用屬性time:intervalDuring連結。	清世宗	http://data.ascdc.tw/time/chi/emporer/001 (如清世宗)	URI	
12	http://data.ascdc.tw/time/chi/emporer/001	中曆生卒年(生年區段、卒	time:intervalDuring	1. 將上一步驟中對應到的"皇帝統治期間"(emperor)class/URI與時間控制表中"朝代"(dynasty)的資訊進行對應。 2. 利用屬性time:intervalDuring連結。	清朝	http://data.ascdc.tw/time/chi/dynasty/001 (如清朝)	URI	
13	http://data.ascdc.tw/dnb/Agent/系統號	西曆生卒年(生年)	schema:birthDate	1. 以"-"進行斷詞。"-"前的數字(組)歸於生年欄位、"-"後的數字(組)歸於卒年欄位 2. 斷詞後資料為"？"、"?"、"[不詳]"者先不轉出。 3. 斷詞、清理後的年號(組)前後有中"["、"]"者先保留並斷開為兩個O值([1788]、[1789])	1733-1818 [1733][1734]-1818	1733 [1733] [1734]	gYear	

Rules for data cleaning, field by field for DNB dataset

LOD Lifecycle- (6)Data Reconciliation

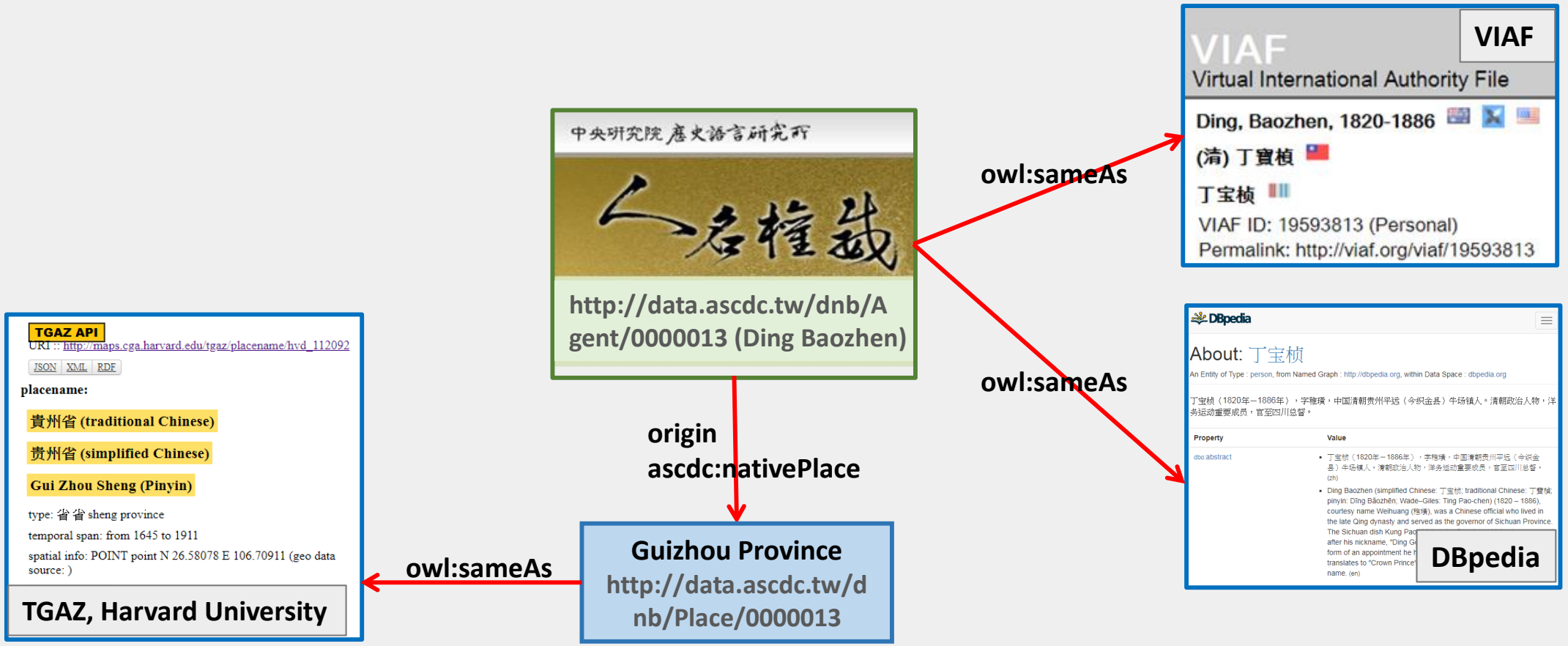
- To design or assign the URIs to each data resource; to reuse and link to the external authority resources; to establish the semantic relations between the resources

URIs design for the DNB LOD dataset

	A	B	C	D
1		**class 類型名稱/ class type**	**URI規劃/ URI Design**	**內容/ Definition**
2		Agent	http://data.ascdc.tw/dnb/Agent/系統號	人名權威檔人物主體實例(主要人物，資料庫中有收錄)
3		Imperial Examination Event	http://data.ascdc.tw/dnb/ImperialExamination/系統號/000001	人物主體相關之科舉出身事件類型實例
4		Official Service Event	http://data.ascdc.tw/dnb/OfficialService/系統號/000001	人物主體相關之任職履歷事件類型實例
5		Name	http://data.ascdc.tw/dnb/Name/系統號/000001	人物主體人名稱謂資訊實例(如仲尼、五柳先生、太白等)
6		Name Type	http://data.ascdc.tw/dnb/NameType/courtesyName	人名稱謂類型實例(如字、號、諡名等)
7		Place	http://data.ascdc.tw/dnb/Place/0000001	人物主體相關地名實例(不分層級)
8		Position	http://data.ascdc.tw/dnb/Position/0000001	人物主體相關之任職履歷事件職位名稱實例(如山西總督、庶吉士)
9		Result Type	http://data.ascdc.tw/dnb/ResultType/001	人物出身科舉事件中獲得科舉等第類型實例(如進士、狀元)
10		Time Interval	http://data.ascdc.tw/dnb/TimeInterval/000000001	人物官職履歷、生卒事件時間區段實例(如開元三年-開元五年、元始四年-居攝三年)
11		Type Access	http://data.ascdc.tw/dnb/TypeAccess/001	人物職官取得之類型實例(如追贈、捐官)
12		Biographical Information	http://data.ascdc.tw/dnb/BiographicalInformation/系統號/001	人物主體略傳資訊實例(人物傳略資訊為主)
13		Work	http://data.ascdc.tw/dnb/Work/系統號/001	人物主體相關之著作作品實例
14		Organization Type	http://data.ascdc.tw/dnb/OrganizationType/001	人物主體官職職位隸屬機構類型實例(如中央機構、地方機構)
15		Organization Hierarchy	http://data.ascdc.tw/dnb/OrganizationHierarchy/001	人物主體官職職位隸屬機構層級類型實例(如一級機構、二級機構)

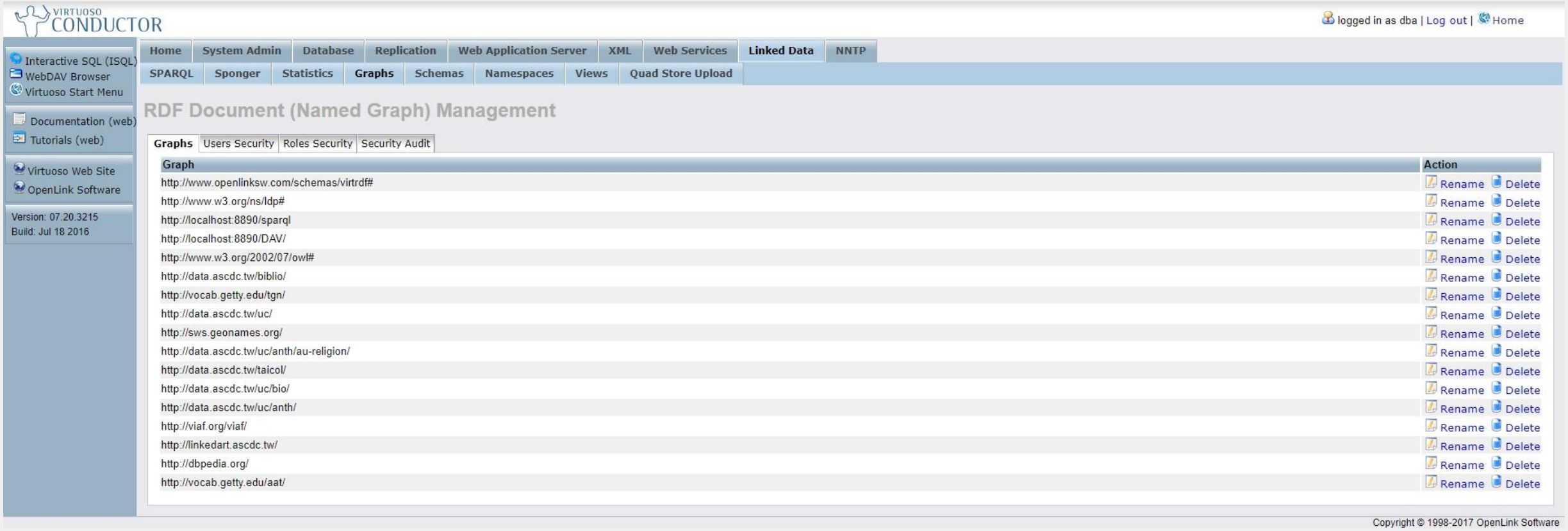
LOD Lifecycle- (7)Data Enrichment

- To create the linkages between heterogeneous resources through the technology of the Name Entity Recognition



LOD Lifecycle- (8)Data Conversion

- To convert, release the data in RDF file format (such as N-Triples, Turtle) and upload into the triple store



Converted LOD data in triple store

LOD Lifecycle- (9)Data Querying in SPARQL

Establish the SPARQL endpoint and offer it with services for semantic queries by examples

SPARQL查詢

Available query types

Category	Query examples
Biology	EX1. Which people in the Qing dynasty came from Fujian Province?
Anthropology	EX2. Which works did Wang Shi-jen (王世貞), a Ming author who had a good command in both poetry and prose, compose?
Religion	EX3. Who took the office of provincial governor (巡撫/ Shunfu) in Anhui Province? How long did they serve?
Name authorities	EX11. How old were the Qing agents when they first took up the post of provincial governor (巡撫/ shunfu)?
	EX12. Where did the Qing agents ranked as jinshi come?

SPARQL query reasoning

1) Set S as the URI of agents, P as "ascdc:dynastyPersonBirth", and O as the URI corresponding to the Qing dynasty.
2) Set S as the URI of agents satisfying the quality defined by step 1, P as "ascdc:imperialExamination", and O as the URI of Posting Event.
3) Set S as the (O) value defined by step 2, P as "leo:involved" and select those whose (O) URI is "jinshi".
4) Set S as the URI of agents satisfying the quality defined by steps 1 – 3 and (O) as the value which is identical to ASCDC's controlled vocabulary of Places and connected to the URI of TGAZ.

```
PREFIX ascdc: <http://data.ascdc.tw/property/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX leo: <http://linkedevents.org/ontology/>
SELECT ?name ?placeCh ?TGAZ_URI
FROM <http://data.ascdc.tw/dnb/Agent/>
WHERE {
    ?systemNumber skos:prefLabel ?name .
    ?systemNumber ascdc:dynastyPersonBirth  <http://data.ascdc.tw/time/chi/dynasty/031> .
    ?imperialExamUri leo:involved  <http://data.ascdc.tw/dnb/ResultType/003> .
    ?systemNumber ascdc:imperialExamination ?imperialExamUri .
    ?placeCtvc owl:sameAs ?TGAZ_URI .
    FILTER contains(STR(?TGAZ_URI), 'tgaz') .
    ?placeCtvc ascdc:nativePlace ?placeCh .
    ?systemNumber ascdc:nativePlace ?placeCtvc .
}
```

SUBMIT

Triples | Map | Chart | Relation chart

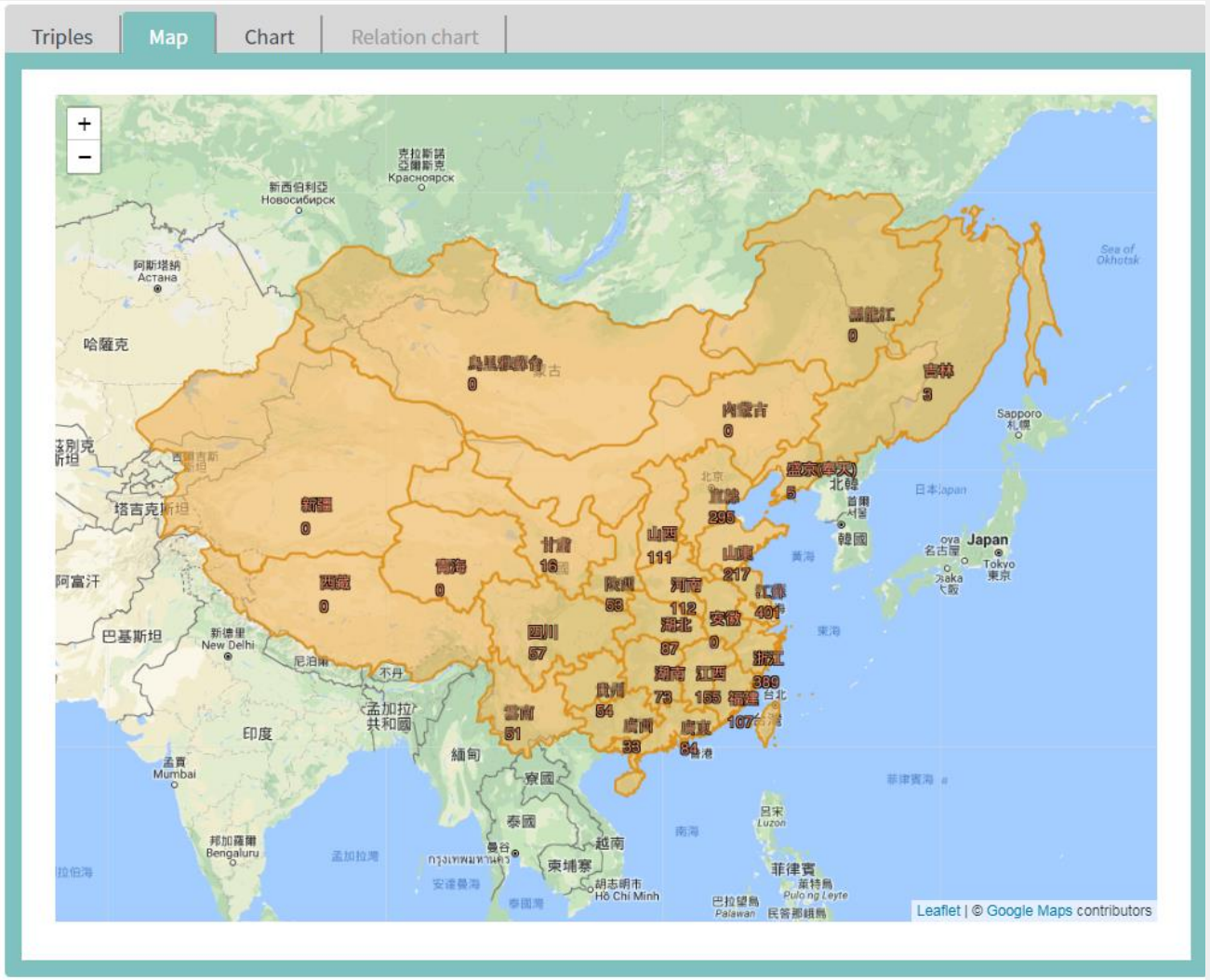
RESULTS ⬇ JSON XML TTL CSV

name	placeCh	TGAZ_URI
鄭敦謹	湖南省	http://maps.cga.harvard.edu/tgaz/placename/hvd_30161
常大淳	湖南省	http://maps.cga.harvard.edu/tgaz/placename/hvd_30161
胡林翼	湖南省	http://maps.cga.harvard.edu/tgaz/placename/hvd_30161
劉權之	湖南省	http://maps.cga.harvard.edu/tgaz/placename/hvd_30161
曾國藩	湖南省	http://maps.cga.harvard.edu/tgaz/placename/hvd_30161
瞿鴻禨	湖南省	http://maps.cga.harvard.edu/tgaz/placename/hvd_30161
羅源漢	湖南省	http://maps.cga.harvard.edu/tgaz/placename/hvd_30161
黎培敬	湖南省	http://maps.cga.harvard.edu/tgaz/placename/hvd_30161
黃倬	湖南省	http://maps.cga.harvard.edu/tgaz/placename/hvd_30161
周有聲	湖南省	http://maps.cga.harvard.edu/tgaz/placename/hvd_30161
陶澍	湖南省	http://maps.cga.harvard.edu/tgaz/placename/hvd_30161
李星沅	湖南省	http://maps.cga.harvard.edu/tgaz/placename/hvd_30161

‹ 1 | 2 | 3 | 4 | 5 | 6 .... 117 ›

SPARQL query examples of DNB datasets

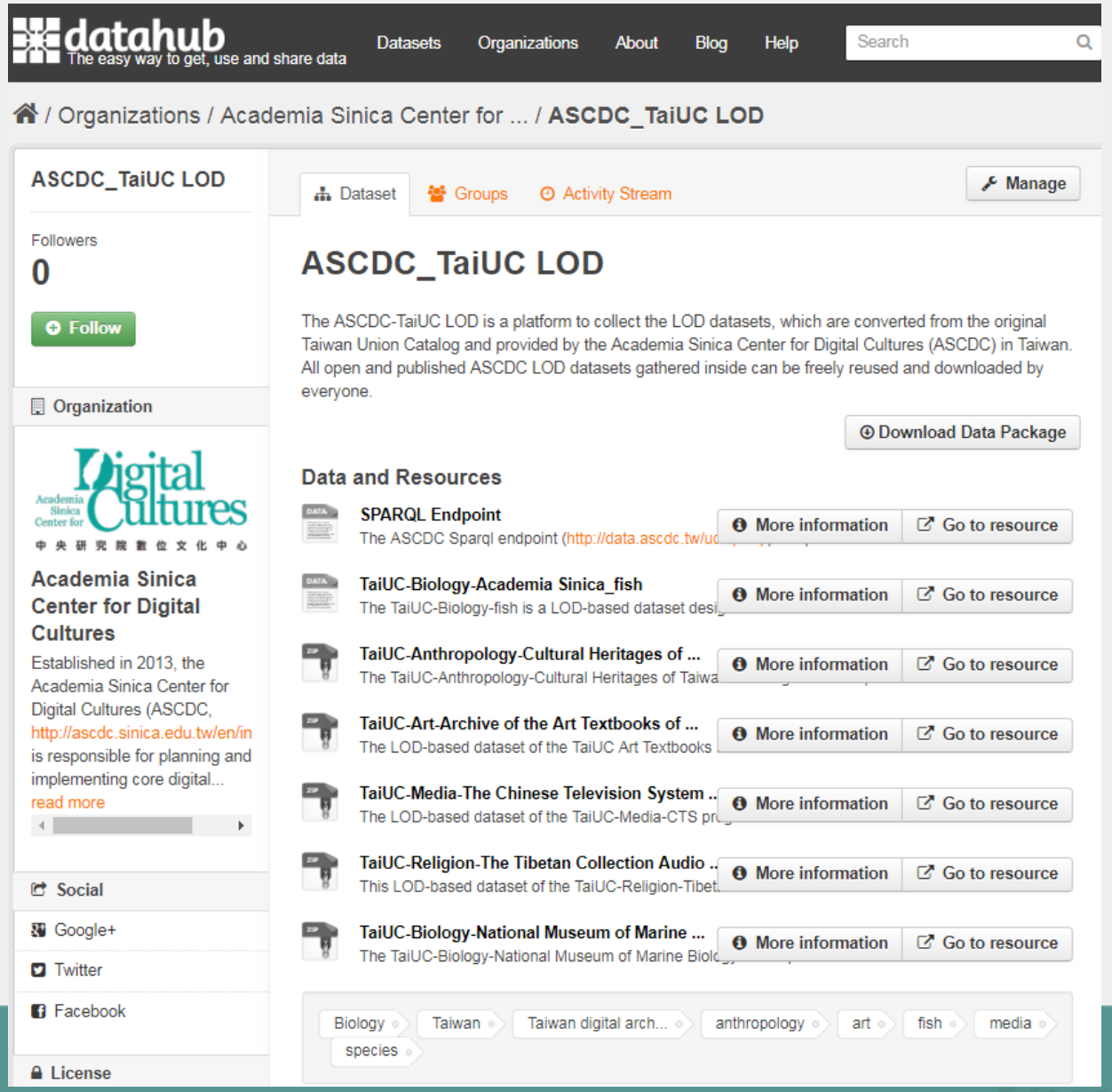
LOD Lifecycle- (9)Data Querying in SPARQL



SPARQL query examples of DNB datasets

LOD Lifecycle- (10)Data Publishing

- To publish the LOD-based datasets in the worldwide open data platform for data sharing and distribution (eg. datahob.io)

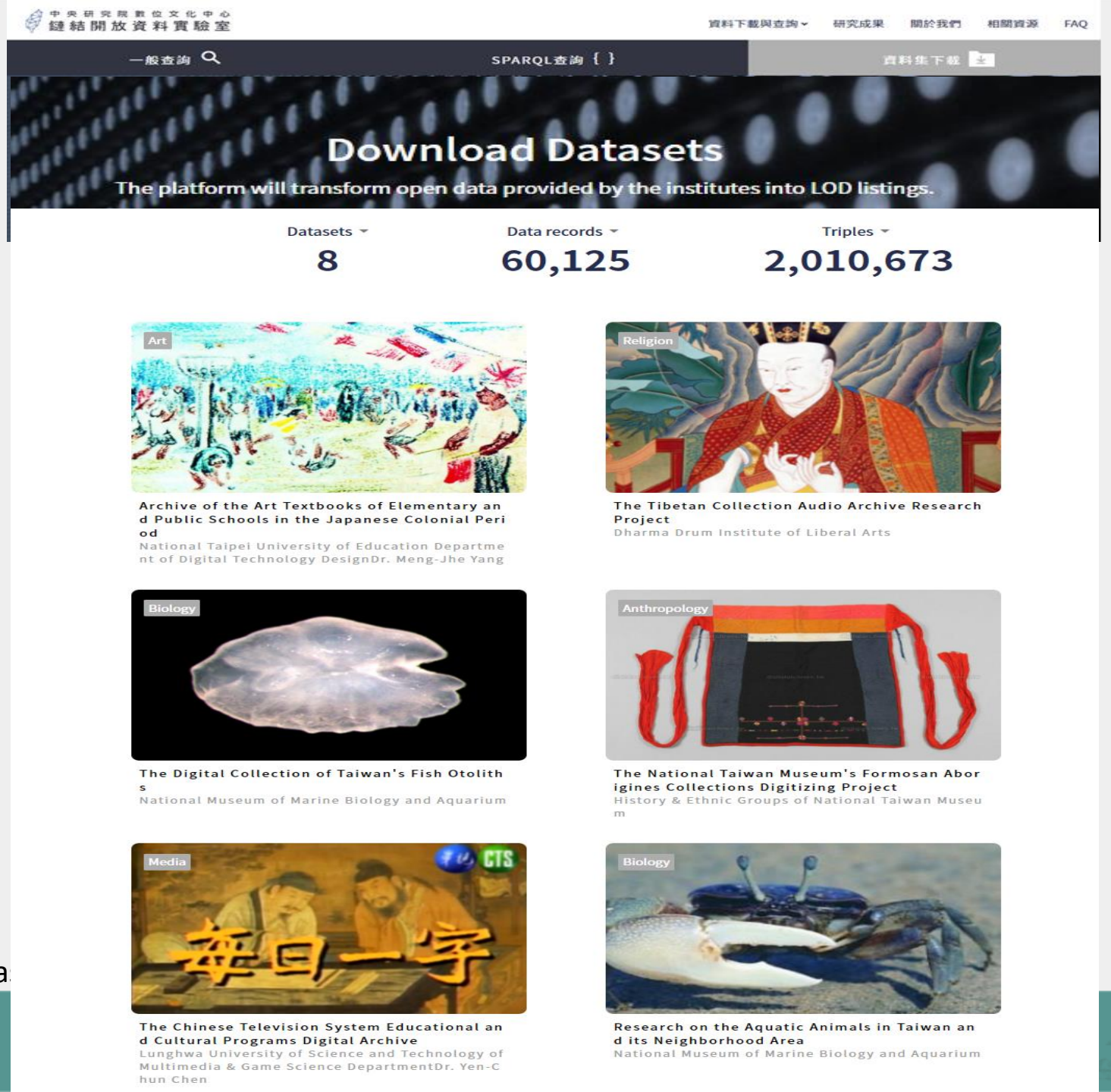


Publishing TaiUC LOD datasets in datahub

LOD Lifecycle-(10)Data Publishing

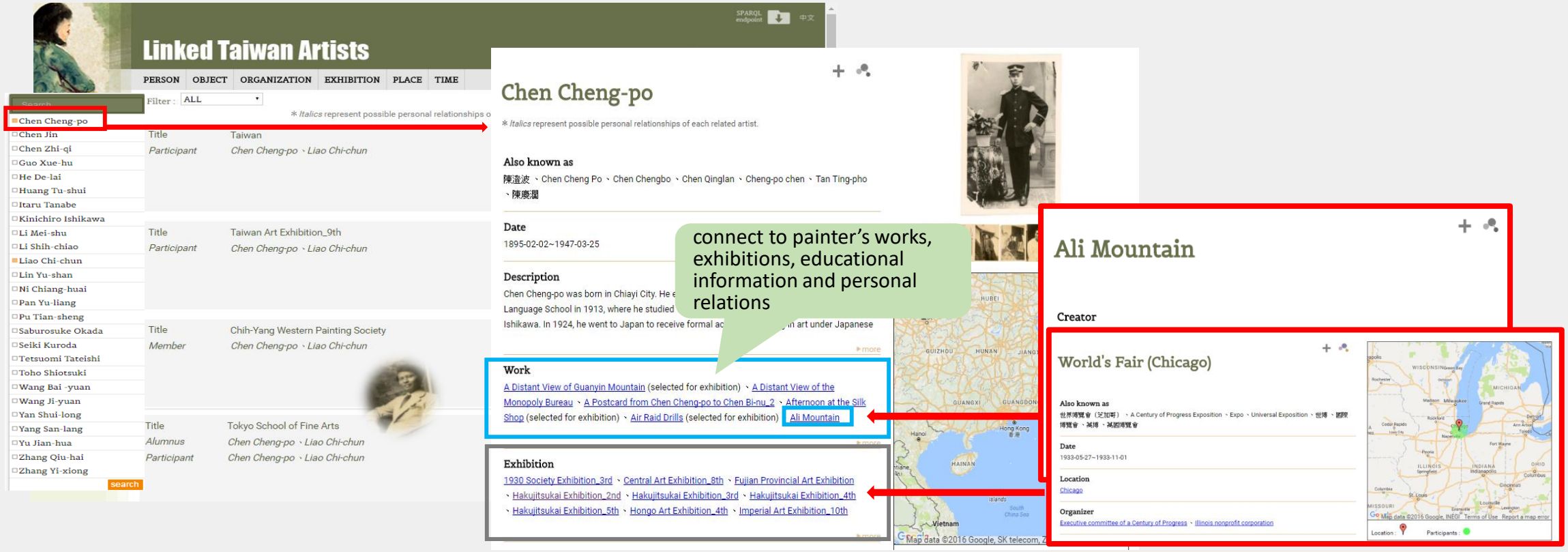
publish the LOD-based datasets in the local-established LOD datasets interface (eg. ASCDC LOD datasets)

Published LOD data



LOD Lifecycle- (11)Application Systems

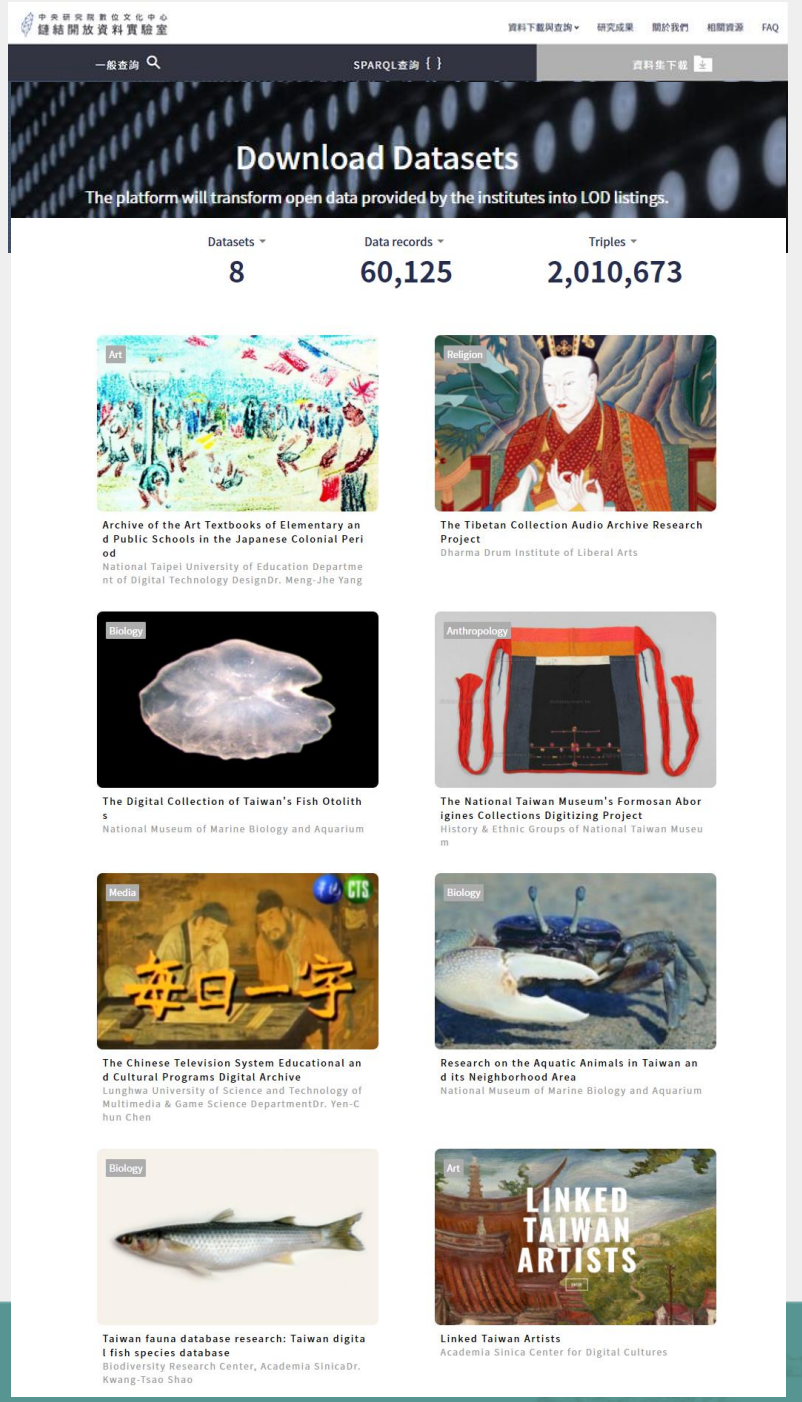
- Develop the application system for providing the LOD-based service system (such as a website)



Linked Taiwan Artists showing information on the early modern painters in Taiwan

LOD in ASCDC: a current view

- 2017: start to execute the LOD programs (conversion of the TaiUC's metadata into LOD-based dataset)
- 2018: establish the LODLab, ASCDC
- Current results
  - Datasets: 8
  - Data records: 60,125
  - Triples: 2,010,673
- Topics of LOD datasets
  - Biology (Taiwan digital fish species database)
  - Anthropology (Indigenous people in Taiwan)
  - Arts (Art Textbooks of Elementary and Public Schools in the Japanese Colonial Period)
  - Religion (Tibetan Collection Audio Archive)
  - Audio-video files (Educational and Cultural Programs from the Chinese Television System)



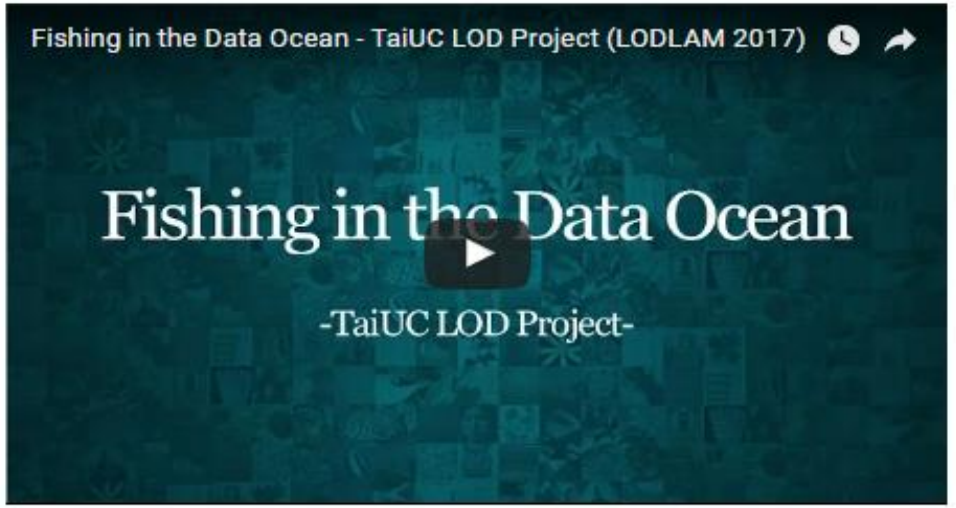
LODLAM Summit 2017

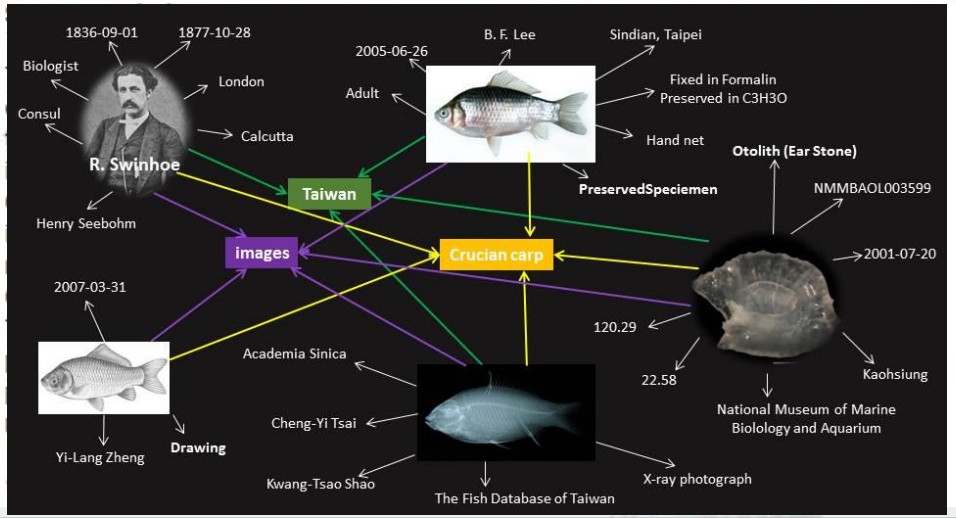
- Top 5 finalists of the Technical Challenge Competition at the 4th Linked Open Data in Libraries, Archives & Museums (LODLAM) Summit in Venice with project “Fishing in the Data Ocean”
- Applying the LOD techniques and standards to bridge Taiwan’s digital heritage with other worldwide relevant biodiversity information on the web.
- Enhancing information retrieval of digital heritages for helping scholars, researchers, general publics and machines to consume data in an open, shareable modus.

FISHING IN THE DATA OCEAN

POSTED BY REDAZIONE | 10 APRIL 2017

Title: Fishing in the Data Ocean
Team: Academia Sinica Center for Digital Cultures (ASCDC)





https://summit2017.lodlam.net/2017/04/12/fishing-in-the-data-ocean/

Data sharing: SPARQL endpoint

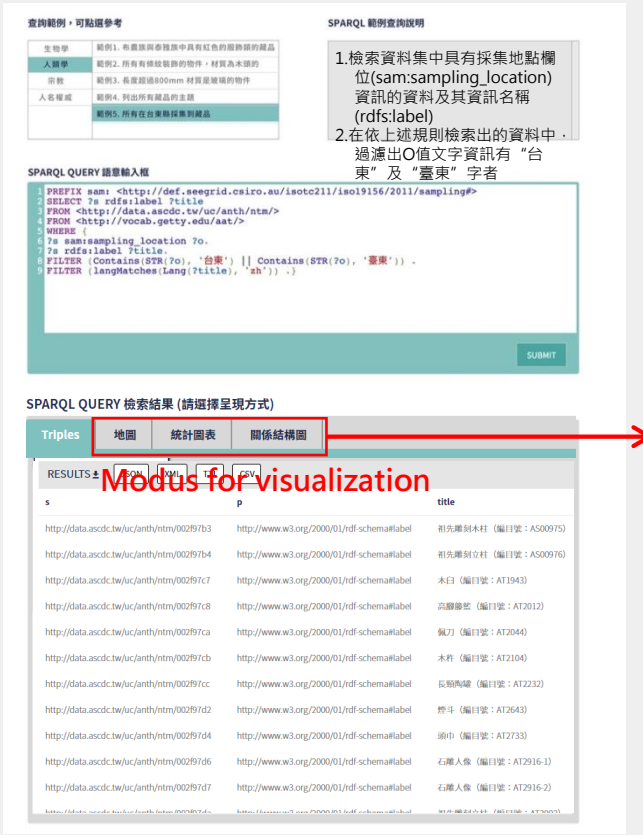
- ASCDC SPAQRL endpoint as platform for SPARQL querying and datasets-downloading
- Each data records and queried results can be downloaded in JSON, XML, TTL and CSV files
- SPARQL query by examples are designed as a user-friendly interface endowing with human readable explanation of how the results of queries are found by machine reasoning

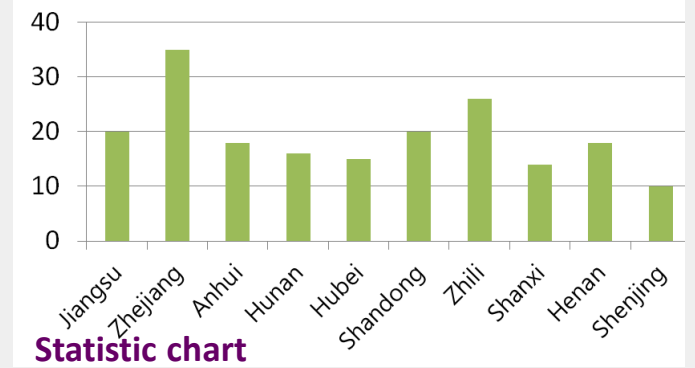
A data record of

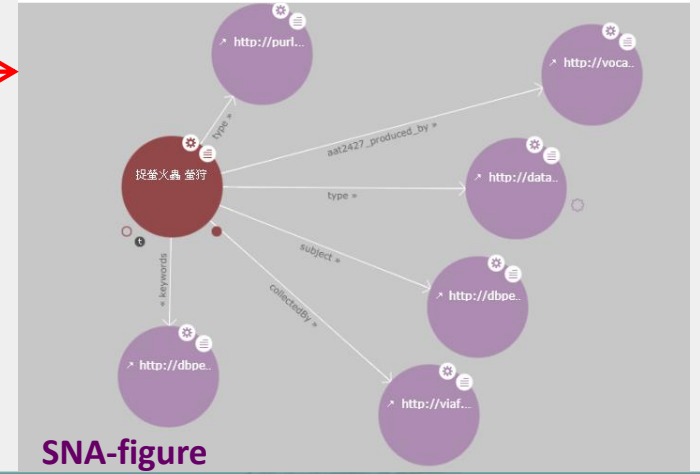
http://data.ascdc.tw/uc/sparql.php

A future: Visualization in LOD

- Visualize the querying results as method to inspire new visions and ideas for the scholar's research and results interpretation
- Applying GIS technique, open source/ API for visualization (such as Google charts…) to create user-friendly access to the data
- Types of visualization in LOD
  - GIS maps
  - Statistic charts
  - Figures of Social Network Analysis (SNA)







PEN museum
開 放 博 物 館

Academia Sinica Center for Digital Cultures

Open Museum

Free platform for digital curation and co-creation

An open museum for everyone

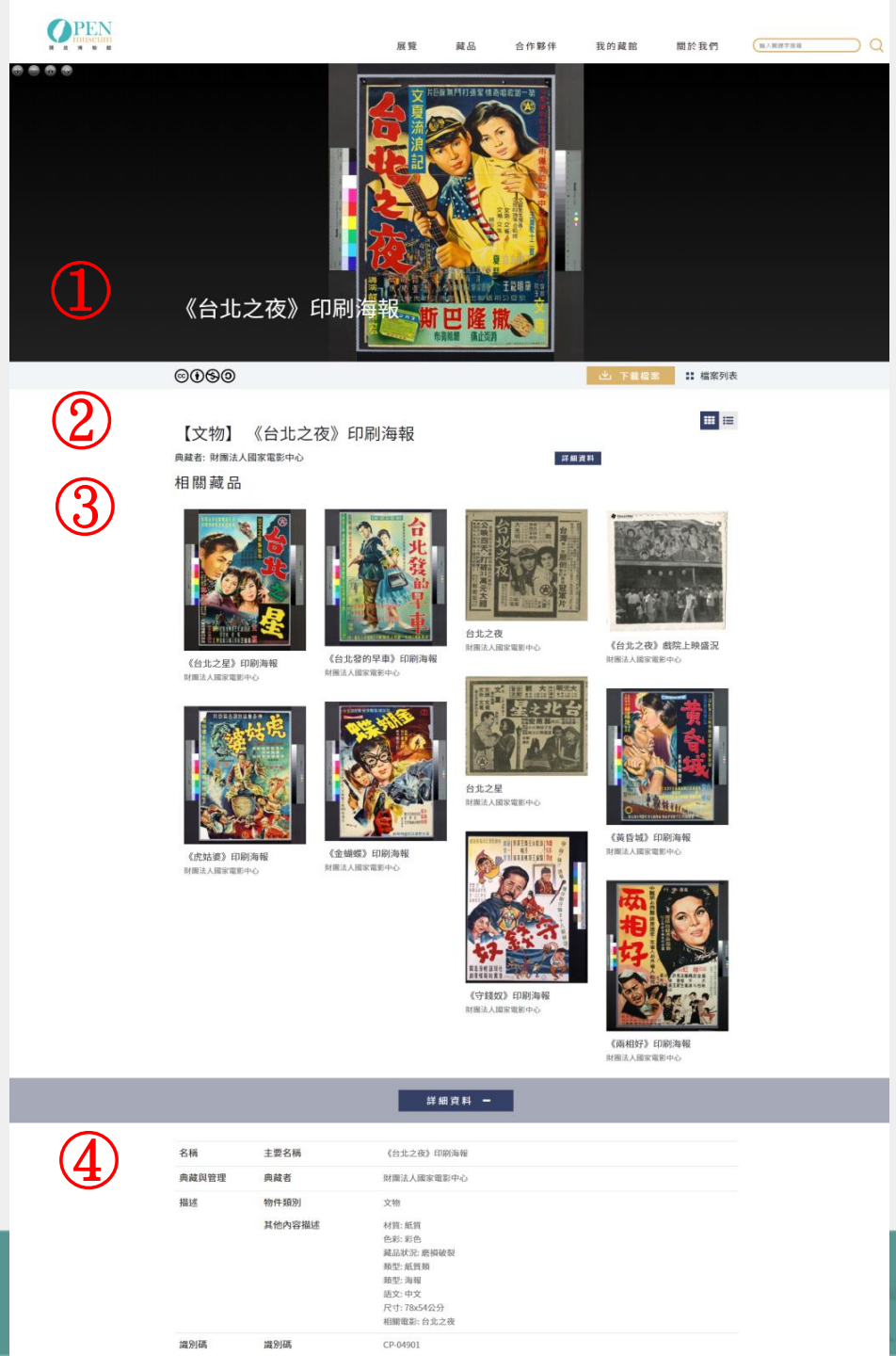
- Open, free space for digital exhibition and co-creation of digital cultural heritages accessible to individuals and institutions
- Users can publish, upload their image collections in different modes for online presentations and share with others for reusing
- Images are opened in CC license
- More than 21,000 resources of image, movies and music recordings, cultural objects are available

http://openmuseum.tw/

Models for online presentations

Type of models and tools:

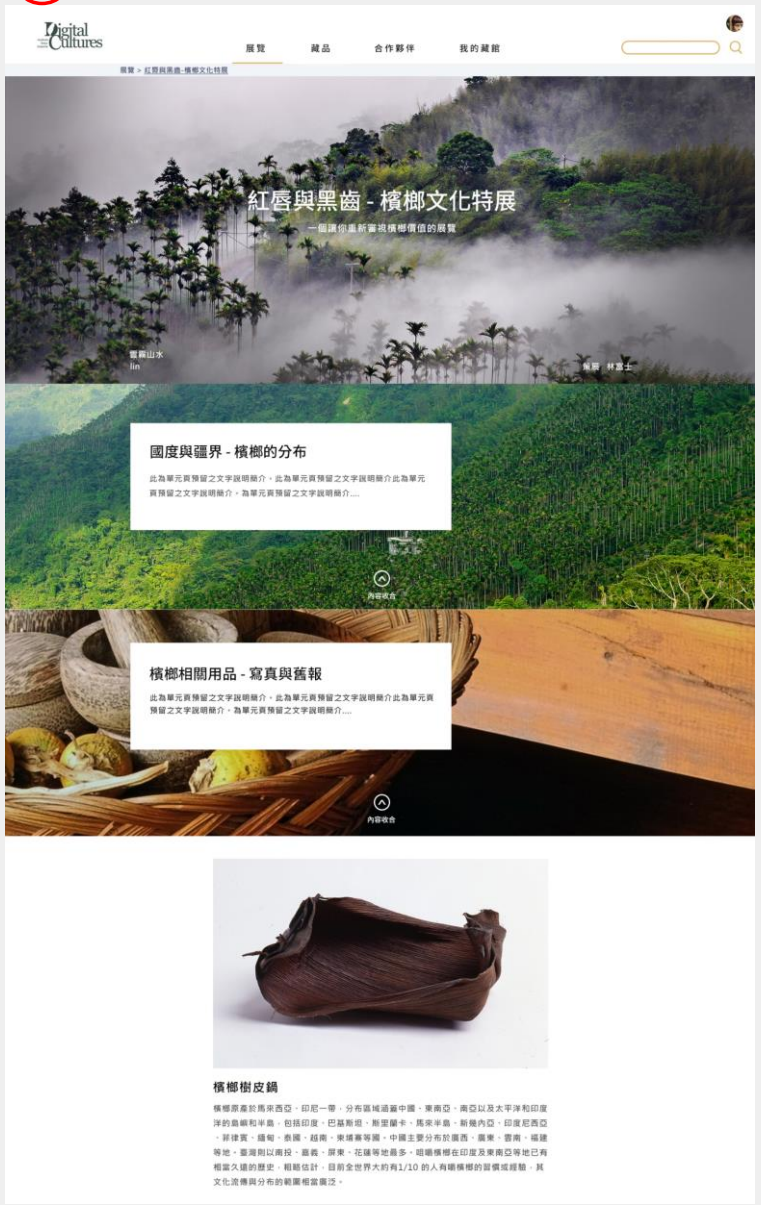
- Modes for online preservation
- Modes for online exhibition
- Tools for online curation



Models for online preservation

1. Image of digitized item
2. Title/ Label of digitized item
3. Resources of related item
4. Metadata of digitized Item

①

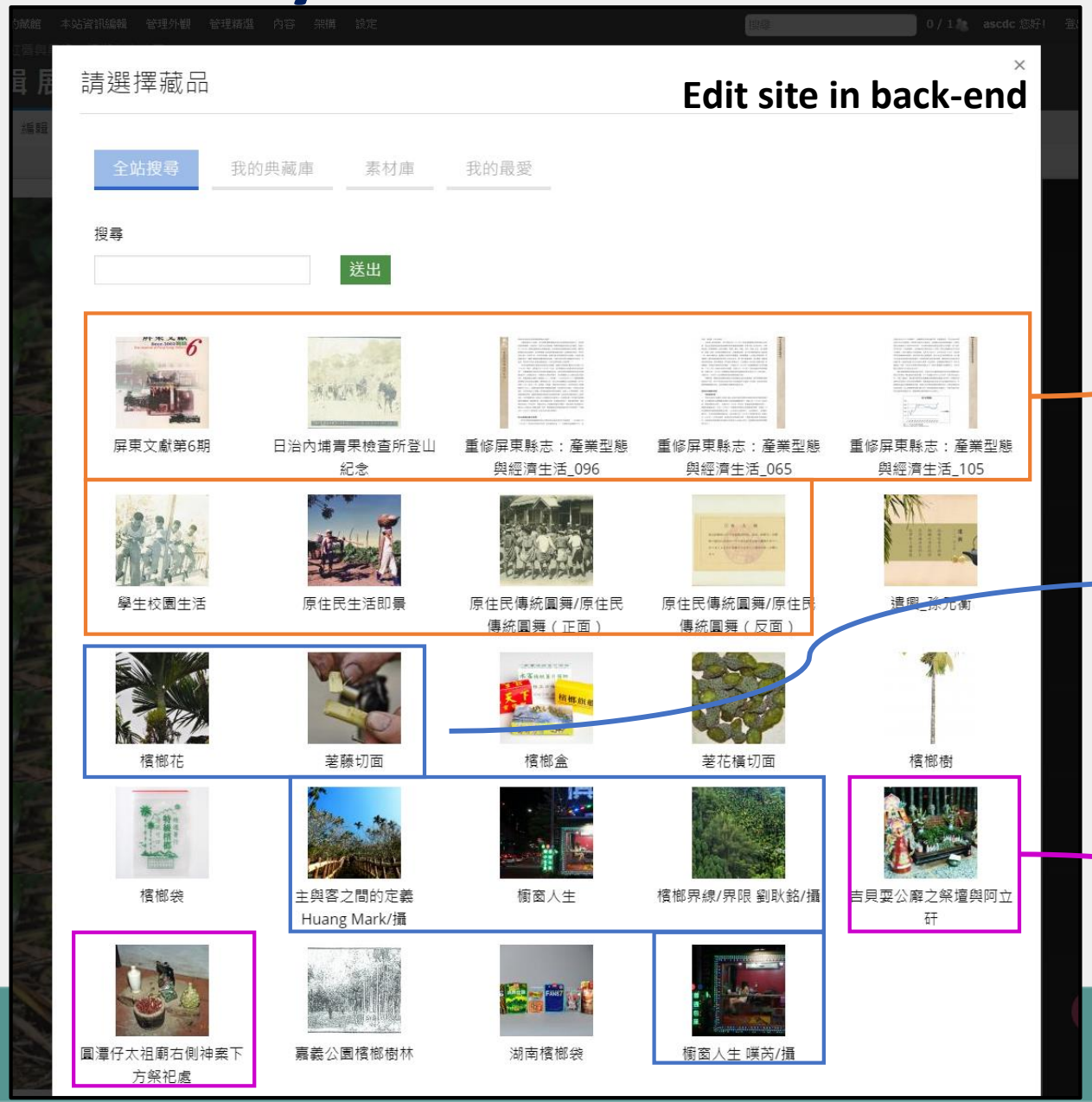


②

Models for online exhibition

1. Vertical presentation of items
2. Horizontal presentation of items

Flexibility to curate exhibition



1. **Users in the editor interface (system back-end) can select images from different sources (cross-source integration)**
2. **More flexible for online curation)**

屏東縣政府文化處

Source A:
Cultural Affairs Department of Pingtung County Government

Source B:
Digital Island/ ASCDC
Academia Sinica

中央研究院 民族學研究所

Source C:
Institute of Ethnology
Academia Sinica

Tools for online curation

Timeline display

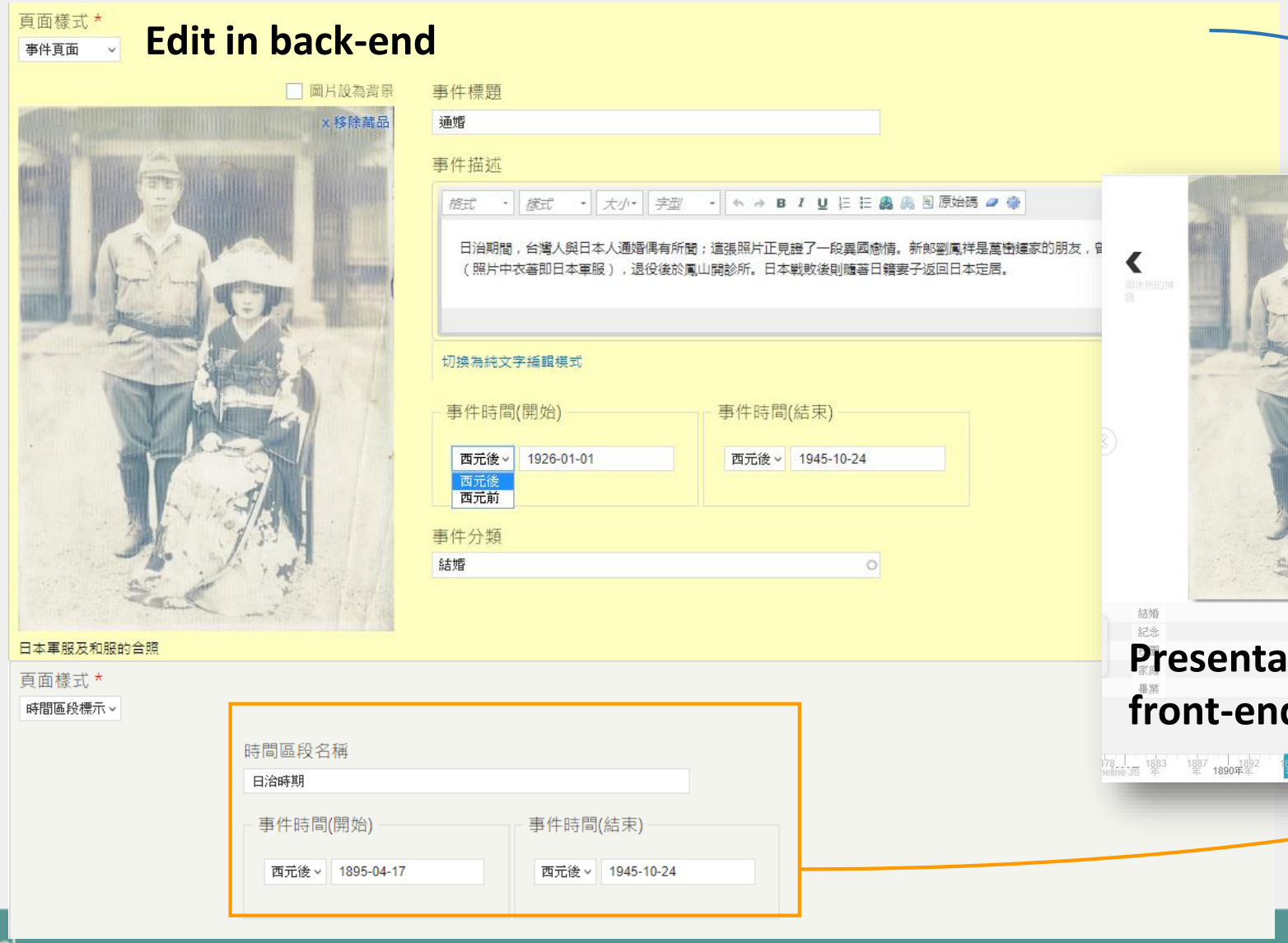
Displaying digitized items or events related to a resource according to chronological sequence

GIS map

Showing the places on map, which related to any information on the digitized object

The Timeline Builder

Edit in back-end

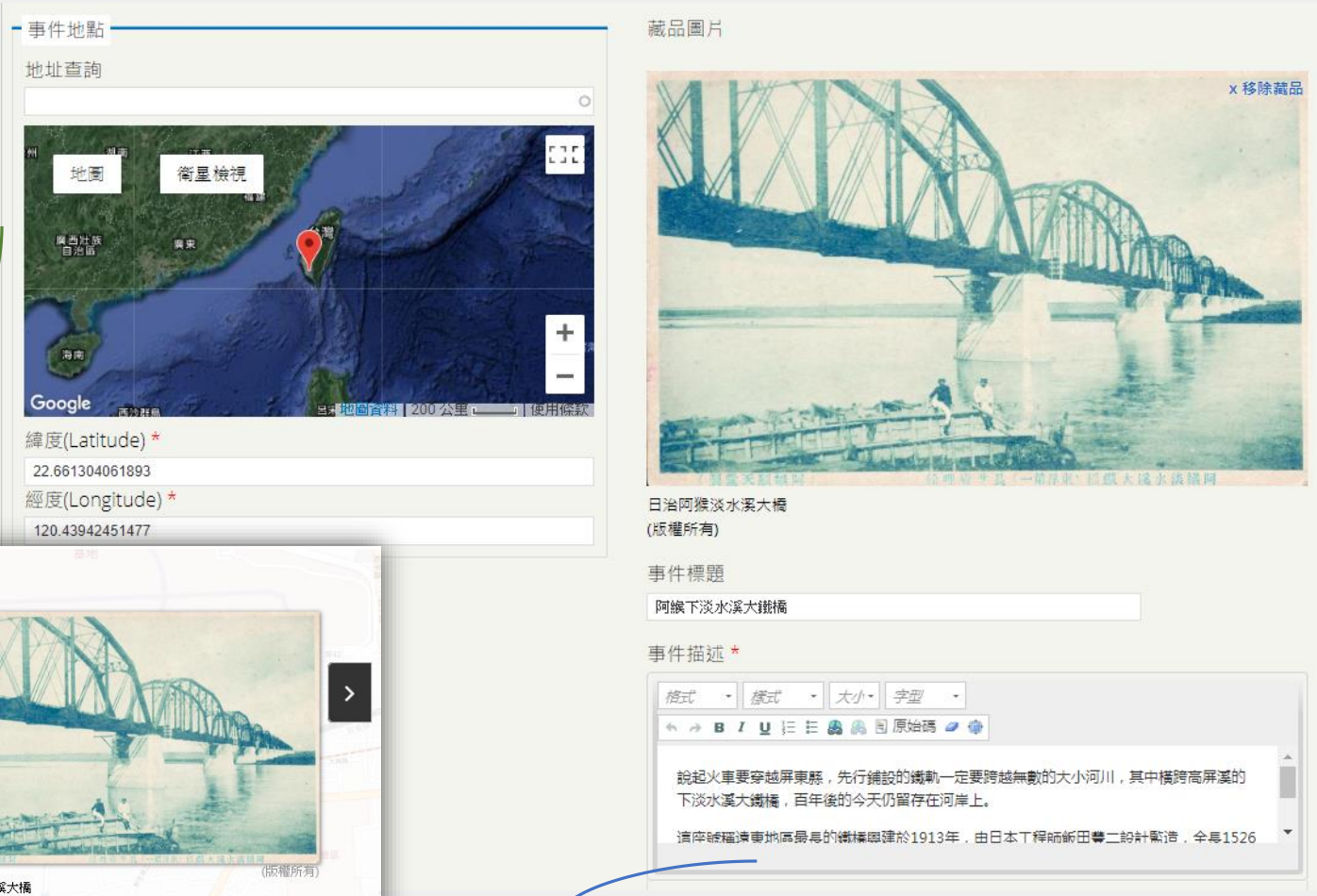


Presentation in front-end

流轉年華 屏東縣老照片專輯：http://pingtung.openmuseum.tw/muse/exhibition/f28f5bd637dfdddab3a5841a4e508810

GIS map: the Story map Builder

Edit in back-end



Presentation in front-end

踏訪阿緱城 屏東縣老照片專輯：
http://pingtung.openmuseum.tw/muse/exhibition/6678f01b9f000f25c474d9c1d7724c6e

Crowdsourcing (in Plan)

This chief is my grandfather!

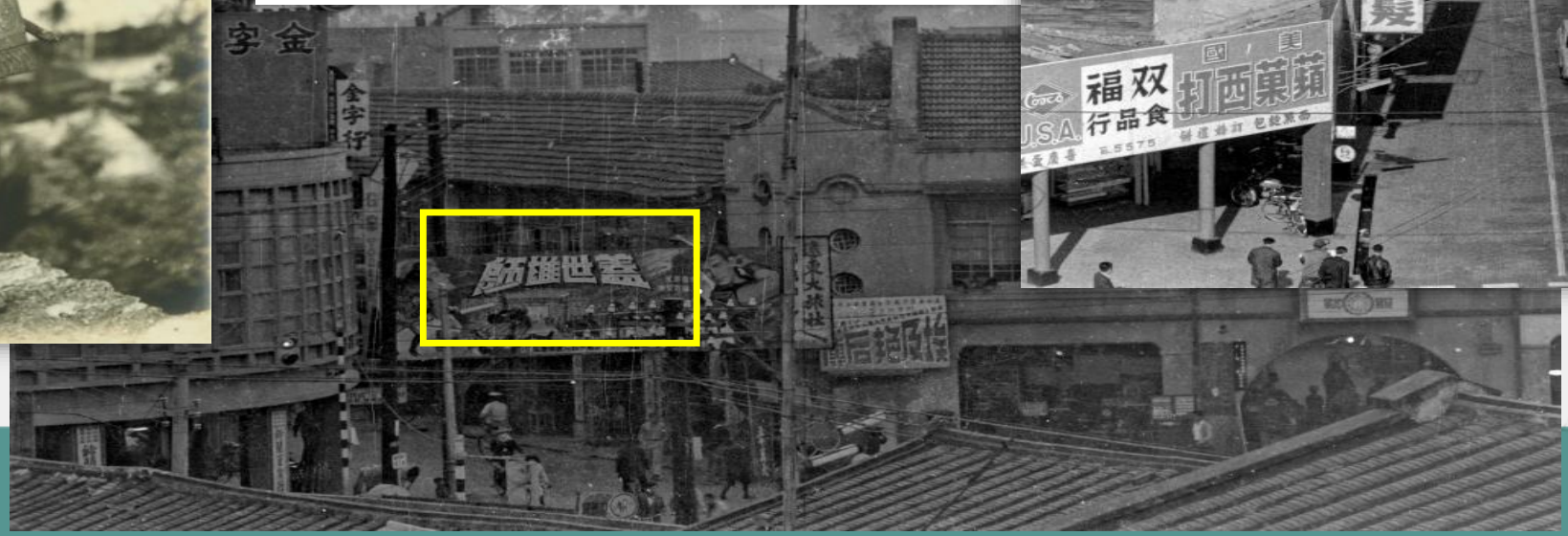
My mom was once lived next to the barber shop 芳美理髮廳！

相關人物 John Doe

攝錄地點 Taichung

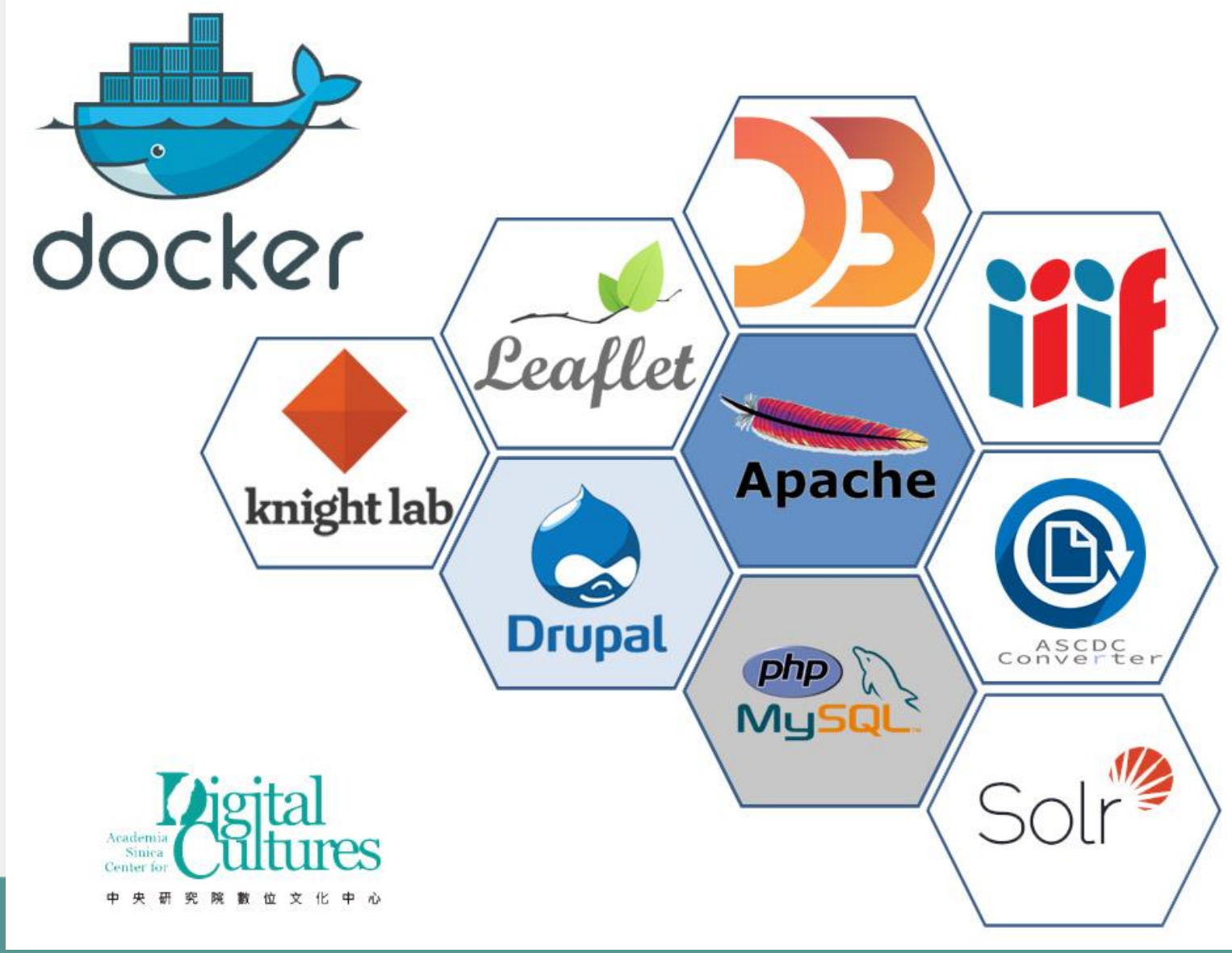
The film《蓋世雄師》is released in 19YY.

攝錄時間 19YY



Future: Image annotation with IIIF

Reusing the open sources for structuring framework



Partners of Open Museum

2017.12

屏東縣政府
Pingtung County Government
243 items

2018.6

國家電影中心
Taiwan Film Institute
764 items

2018.10

財團法人公共電視文化企業基金會
Public Television Service Foundation
3,000 items

中研院生物多樣性研究中心
Biodiversity Research Center, Academia Sinica
ca.3,000 items

中研院史語所歷史文物陳列館
Museum of the Institute of History and Philology, Academia Sinica
131 items

2018.12

臺灣音樂館
Taiwan Music Institute
12,128 items

余如季攝影作品
Photographical works of Yu JU-Chi
2,532 items

鄧雨賢音樂藏品
Music recordings of Teng Yu-Hsien
307 items

Academia Sinica Center for Digital Cultures

Creative Comic Collection (CCC)

Innovative application of digital contents

CCC project: creative re-using of digital content

- Goal: Turning the cultural heritages/ results of academic researches into re-usable intellectual and artistic assets
- Promoting the knowledge of DH in a more interesting, easily understandable way to general publics
- How to do?

– CCC comic book series
– Walking into the past and present series
– Exhibitions and workshops

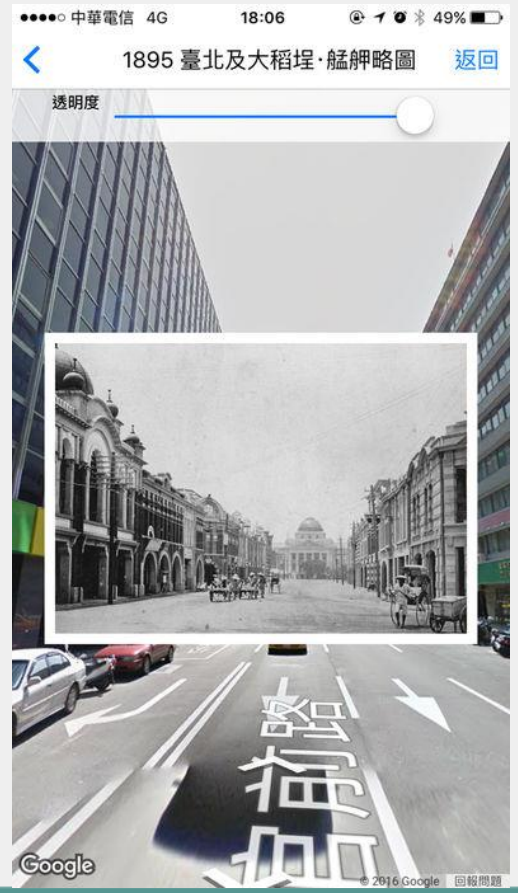
https://www.facebook.com/CreativeComicCollection/

Creative Comic Collection

- Manga series collaborated with Ministry of Culture, Taiwan since 2009 and reorganized in 2017
- Inviting comics artists to develop creative topics based on scholars' study results in DH
- A new model for creative industry to reuse the intellectual property and enhancing communication between the academic institutions and general publics

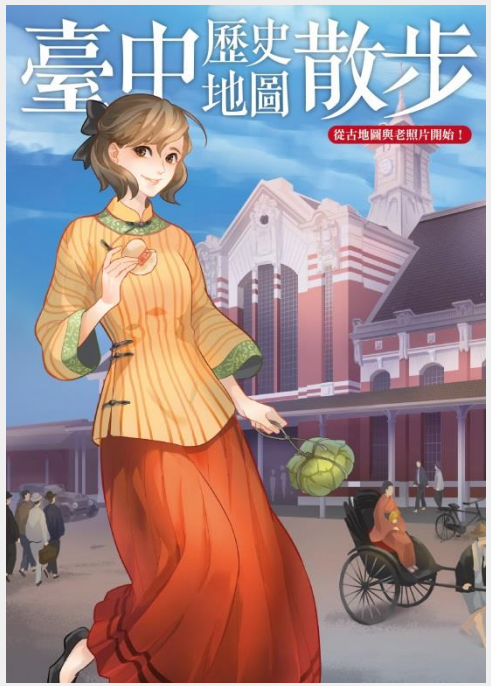
Walking into the past and present series

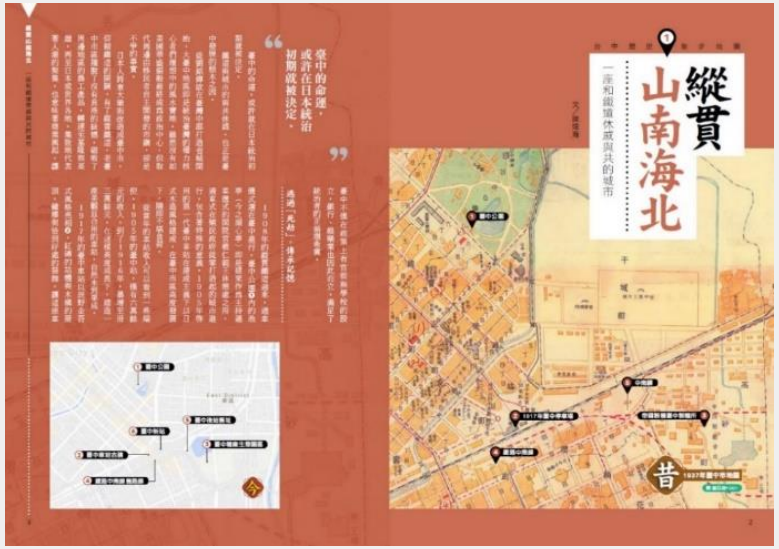
- Started from 2016, book series of Taiwan's city travel guide equipped with mobile AAP
- City guides on Taipei, Taichung are available; Tainan is upcoming
- Reusing historical maps and photos from archives in combination with technology of GIS positioning and showing the comparison of a street view on the google map between past and present
- Using finger tips to see the past city views in Taiwan just on the street



Walking into the past and present series

- Guide book: containing essays and illustrations on the cities' history according to scholar's researches
- Forthcoming: English and Japanese translation version

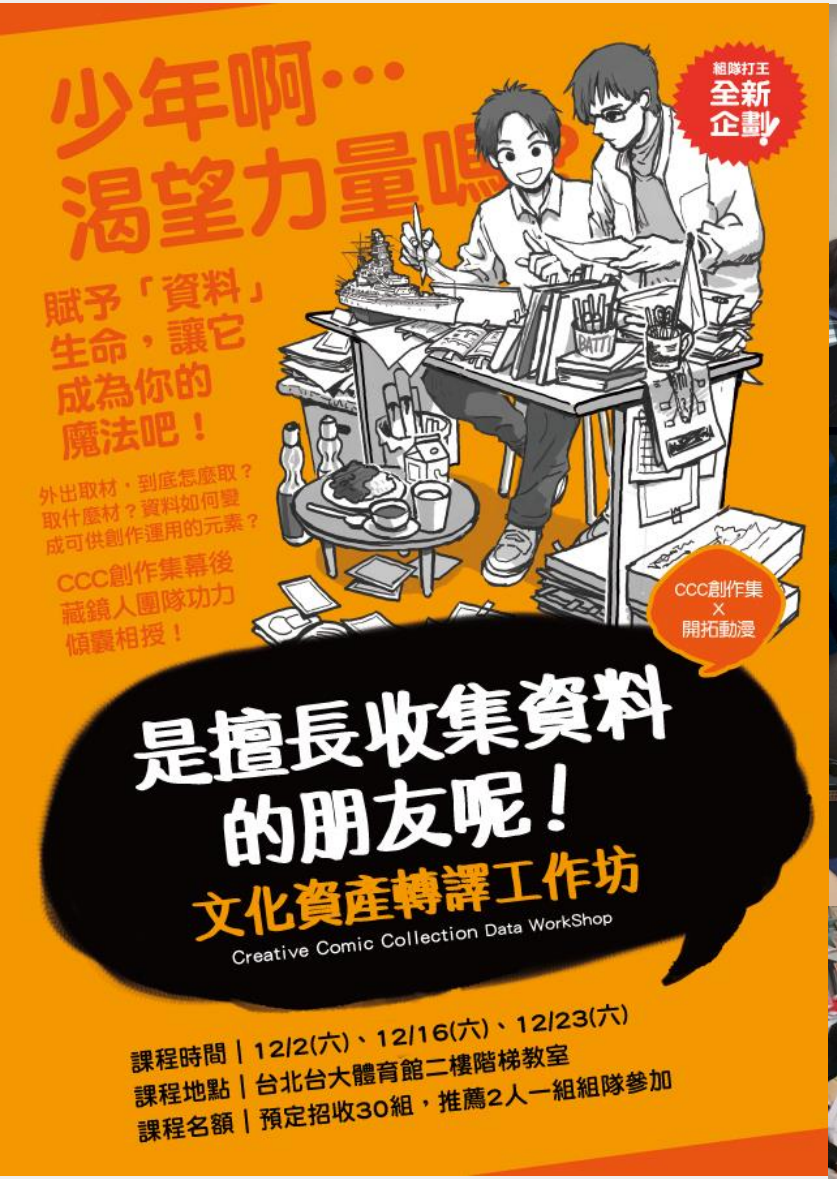




Exhibitions and workshops

- Exhibitions and workshops as platform to promote ASCDC's achievement on the creative reusing of digital content

Exhibition of CCC comic series and AR digital experience
2017.08.15 - 09.17 in National Taiwan Museum, Taipei



Creative comic collection data workshop
Dec. 2, 16, 23, 2017

Data fair workshop
Feb. 10-11, 2018

Education of data reusing- Workshop on creative translation of data

Creative data translation- CCC comic series

Integration of DH content- exhibition with digital application

Academia Sinica Center for Digital Cultures

A 5F, Research Building, Institute of History and Philology
No.128, Sec. 2, Academia Rd., Nangang Dist., Taipei City 11529, Taiwan (R.O.C.)

E ascdc@sinica.edu.tw

P +886-02-2652-1885

F http://ascdc.sinica.edu.tw/en/

Q & A

ASCDC

Digital Cultures
中央研究院數位文化中心

THANK YOU !