

數位人文環境下BIBFRAME書目框架在 古漢籍善本鏈結資料的再用與擴展

陳淑君 (Sophy Shu-Jiun Chen)

中央研究院歷史語言研究所/中央研究院數位文化中心

Institute of History and Philology / Academia Sinica Center for Digital Cultures


2021 年第 13 次中文文獻資源共建共用合作會議：【跨域經緯。創新翻轉圖書館與數位人文的對話】
2021年 7 月 13-14 日 線上視訊會議

研究背景

- **BIBFRAME**為美國國會圖書館發展的**書目知識本體**，用以**取代圖書館社群的機讀編目格式 (MARC) 標準**。自BIBFRAME於2012年公布後 (2012 V. 1.0、2016 V. 2.0)，在圖書館專業與學術社群已有諸多研究實驗，例如：倫敦大學學院資訊學系發展基於BIBFRAME的鏈結開放資料書目資料集 (鏈結開放書目資料)、Linked Data for Production (簡稱LD4P) 研究計畫聯合包括美國國會圖書館及史丹佛大學在內等六個機構共同合作，將圖書館的技術服務生產工作流程，從傳統轉換為以鏈結開放資料為基礎的方式
- 本研究以**古漢籍善本書目資料**為對象，採用**BIBFRAME 2.0**作為知識本體進行**鏈結資料 (Linked Data) 轉置基礎**，探究如何運用**語意檢索 (SPARQL)**、**分析與視覺化工具**進行人文學者在不同研究階段的**問題探究**
- **研究問題：**
 - (一) BIBFRAME作為書目知識本體，滿足古漢籍善本書目的程度為何? 對於無法滿足的部分，如何擴充以維持BIBFRAME架構的一致性?
 - (二) 古漢籍善本書目的鏈結開放資料集如何支援數位人文研究?

研究對象

A Digital Library of Chinese Rare Books



古漢籍善本
數位化資料庫
檢索系統

進階檢索 主題瀏覽 下載專區

本計畫由蔣經國國際學術交流基金會委託贊助
資料庫內容為中央研究院歷史語言研究所傅斯年圖書館、美國國會圖書館亞洲部、哈佛大學哈佛燕京圖書館、普林斯頓大學東亞圖書館共同建置
中央研究院歷史語言研究所傅斯年圖書館與資訊科學研究所數位典藏技術發展組共同開發
凡引用本資料庫所有內容，請註明出處 All Rights Reserved

- 中研院傅斯年圖書館2007-2009年，受蔣經國國際學術交流基金會委託，與美國國會圖書館亞洲部（Library of Congress · Asian Division）、普林斯頓大學東亞圖書館（East Asian Library of the Princeton University）、哈佛大學燕京圖書館（Harvard-Yenching Library of the Harvard University）合作建置的「古漢籍善本數位化資料庫」精選書目藏品
- 內容：13部集叢（series）、1,070部單行本（monographies）、6,352部子目（volumes），總計7,435筆階層書目資料，書目資訊涵蓋明朝初年至二十世紀中葉（1368-1953）近六百年時間（包含印刷刊本、手抄本）。

「古漢籍善本資料庫」善本藏品來源、重點典藏書目分類概覽

館藏單位		經部	史部	子部	集部	總計
中研院傅斯年圖書館	數量	0	130	2	76	208
	重點		地理類		明人詩文集	
美國國會圖書館	數量	21	0	2	273	296
	重點	四書、易經			清人詩文集	
哈佛大學燕京圖書館	數量	28	105	82	121	316
	重點	四書、易經	地理、政書	術數、釋家	清人詩文集	
普林斯頓大學圖書館 東亞部	數量	40	12	101	77	230
	重點	易經、春秋		醫家	清人詩文集	
總計		89	247	187	547	1070

古漢籍善本數位化資料庫：藏品書目分析特色

- 「**學業書籍**」（**科學用書**）**豐富**：坊刻科學用書在美國國會圖書館、哈佛燕京圖書館、普林斯頓大學東亞圖書館清單中均出現，歸納這些書籍可得出九類：「四書類」、「五經類」、「八股類」、「古文選本」、「二三場試墨與範文彙編」、「翰林館課」、「通史類」、「類書」、「諸子彙編」。
- 「**名家評選本**」蔚為風潮：書單中「○○○評註○○詩選」、「○○○評註○○文選」題名甚多，例如《新刻錢太史評註李于鱗唐詩選玉》、《唐會元精選批點唐宋名賢策論文粹》等，顯示明清時期出版常以「名家」之名，標舉所刊刻書籍之精善，既是品質保證，亦是行銷手法。

美國國會圖書館數位化影幅檔案編碼清單(295部)

編號	書名	冊數	索書號
1	七經偶記12卷3冊與四一居士文鈔合1函	3	A101.85W19
2	四一居士文鈔6卷附錄1卷2冊與七經偶記合1函	2	A101.85W19_1
3	五經類語	8	A104L64
4	讀易舉例35卷6冊1函	6	A132.85Y92
5	易通	4	A133.78T19
6	書經體注圖攷大全 據書名葉 6卷卷首1卷4冊1函	4	A142.85F21.1
7	尚書竊8卷4冊1函	4	A143.78H86
8	詩經肯鳳詳解	8	A151.283C42
9	詩故考異32卷8冊2函	8	A151.385H86
10	詳增經旨音釋毛詩白文4卷6冊1函	6	A151.74C11
11	禮經類編30卷10冊1函	10	A160.775L62
12	讀春秋左氏贊言12卷6冊1夾板	6	A171.276W18
13	古邠詩義1卷1冊1函	1	A151.386H85
14	四書圖考13卷12冊1函	12	A190.5T79
15	論語詩箋1卷1冊1函	1	A191.382Y91
16	鄉黨義考7卷7冊1函	7	A191.384H86
17	讀孟子劉記1卷2冊1函	2	A192.384T78
18	四書近見錄 一名筆洞生近見 4卷4冊1函	4	A195.376H85
19	四書大意存11卷10冊1函	10	A195.378M69
20	字學三書15卷5冊1函	5	A240.8T86
21	南行載筆不分卷4冊1函	4	K281L183
22	春鶴堂集17卷4冊1函	4	K282C251

美國國會圖書館清單中舉業經書類善本書籍概覽

古漢籍善本數位化資料庫：藏品書目分析特色

- 「**實用書籍**」**逐漸興盛**：包含「命理書籍」、「醫藥書籍」、「地理方志」與少數「宗教書籍」，顯示這些書籍在明清出版市場中，應當佔有一定比例的銷售量，反映明清出版刊刻事業具商業性、通俗性、實用性之特色。
- 「**文娛小品**」**出版漸增**：書單出現屬於「小品文」等文人閒暇娛樂之作，這類作品專講文人琴棋書畫、蒔花藝草等閒情百態，是文學觀念在晚明時期的一大轉變，亦可從書籍清單中發現出版文化風向也隨之改變，例如《翠娛閣評選鍾伯敬先生小品二卷》。

普林斯敦大學東亞圖書館掃描清單

序號	題名	冊數	Call Number
1	凝園讀易管見：十卷	10	A11/1285
2	先天易貫：五卷	8	A11/2673
3	易翼述信：十二卷	12	A11/2748
4	春秋左傳杜林滙參：三十卷，卷首	16	A101/2957
5	左氏條貫：十八卷	8	A101/3266
6	痘疹類編釋意全書	6	C83/743
7	痘疹定論：四卷	4	C83/747
8	纂修醫學入門：六卷	12	C83/1846
9	羅氏會約醫鏡：三十卷	20	C83/1847
10	便元集：二卷	6	C83/2867
11	成方切用：十二卷，卷首，卷末	20	C83/3497
12	喻氏醫書三種：十六卷	22	C83/3828
13	傷寒第一書：四卷，附錄 [二卷]	6	C85/2909
14	種痘新書：十二卷	6	C86/2542
15	救偏瑣言：十卷	6	C86/2547
16	天花精言：六卷	4	C86/2938
17	痘疹傳心錄：十九卷	12	C86/3054
18	喉科指掌：六卷	2	C88/2868
19	推拿秘書：五卷	2	C91/2538
20	胎產輯萃：四卷	8	C98/1833
21	本草詩箋：十卷	8	C103/1472
22	溫熱暑疫全書：四卷	4	C110/3048
23	補瓢存稿：六卷	6	D33/1182

普林斯頓大學圖書館清單中醫藥類善本書籍概覽

研究步驟

1. 資料授權 Data Licensing

確認與取得資料開放授權條款

2. 資料剖析 Data Profiling

資料結構、知識內涵、權威詞彙分析

3. 資料建模 Data Modeling

知識本體設計及對應分析(BIBFRAME, dbpedia)

4. 資料清理 Data Cleaning

進行結構化資料處理，修正錯誤資料，使資料質量最大化

5. 資料加值 Data Enrichment

從字串到事物，連結外部資源

6. 資料轉置 Data Conversion

轉換資料為RDF(N-Triples /Turtle)

7. 資料查詢 Data Access

建立語意檢索平台 (SPARQL endpoint)，提供語意檢索功能及查詢範例 (query examples)

8. 資料發佈 Data Publishing

將資料集同步發布於全球資料集散平台(datahub.io)或自建之LOD資料集入口介面中、符合五顆星等級LOD

9. 資料應用 Data Consuming

呈現以LOD為基礎的服務系統/提供語意檢索功能及後續在分析工具的應用)

研究結果

古漢籍善本書目鏈結開放資料發佈

- 將資料集、語意模組架構及元素等說明文件同步發布於全球資料集散平台(如 datahub.io)及自建的LOD資料集入口介面。
- 藉此達成資料開放、近用、下載之目標，俾利使用者詳細理解資料範圍、內容、架構及檢索操作。

The screenshot shows the Datahub website interface. The main content area displays the 'Database of Chinese Rare Books (CRB)' dataset page. The page includes a search bar, navigation tabs for 'Dataset', 'Groups', and 'Activity Stream', and a 'Download Data Package' button. The dataset description states: 'The LOD dataset of ASCDC_Database of Chinese Rare Books (CRB) is created by the Academia Sinica Center for Digital Cultures (ASCDC) and based on the metadata from the online database of "a Digital Library of Chinese Rare Books" (古漢籍善本數位化資料庫, <http://rarebookdl.lhp.sinica.edu.tw/rarebook/Search/index.jsp>). This digital library is established and maintained by the Fu Ssu-nien Library at the Institute of History and Philology, Academia Sinica, Taipei, Taiwan (ROC). The data includes titles and information on more than 900 Chinese rare books and manuscripts collected by the Library of Congress, Princeton University Library, Harvard-Yenching Library and Fu Ssu-nien Library. Containing 39,911 data records and over 140,638 triples, the dataset is downloadable in ttl file format. The textual content of the LOD dataset is licensed with the Creative Commons Attribution (CC0).

The 'Data and Resources' section shows the 'LOD dataset of CRB' with a 'More information' button and a 'Go to resource' button. The breadcrumb trail is: Academia Sinica > Chinese rare books > East Asian Library > Fu Ssu-nien Library > Harvard University > Harvard-Yenching LI... > Institute of Histor... > Princeton University > the Asian Division ...

The 'Additional Info' table is as follows:

Field	Value
Author	Academia Sinica Center for Digital Cultures (ASCDC), Taiwan (R.O.C.)

The screenshot shows the Academia Sinica Center for Digital Cultures website. The main content area displays the 'Download Datasets' page. The page includes a search bar, navigation tabs for 'Datasets', 'Data records', and 'Triples', and a 'Download Data Package' button. The dataset statistics are: 10 Datasets, 135,287 Data records, and 4,391,459 Triples.

The 'Download Datasets' section shows three datasets:

- Biology**: Taiwan fauna database research: Taiwan digital fish species database. Biodiversity Research Center, Academia Sinica Dr. K wang Tsao Shiao.
- Art**: Linked Taiwan Artists. Academia Sinica Center for Digital Cultures.
- History**: Linked Data for Name Authority Database at IHP. Institute of History and Philology, Academia Sinica Dr. Cheng-Yun Liu.

The 'Constructing LOD (Linked Open Data) for Chinese Rare Books' section shows the Fu Ssu-nien Library of the Institute of History and Philology, Academia Sinica.

發佈古漢籍善本LOD資料集於Datahub (左)及中研院LOD Dataset資料平臺介面(右，綠色框線)，以利資源開放、近用

研究結果

古漢籍善本知識本體：BIBFRAME的再用與擴展

- 本研究以2016年由美國國會圖書館（Library of Congress）發布的「書目框架」（Bibliographic Framework, BIBFRAME）2.0版本為基礎，建構「古漢籍善本知識本體」（Chinese Rare Book Ontology, 以下簡稱CRBO），分別由**26個「類別」（Classes）**及**89個「屬性」（Properties）**組成，包括作品（Work）、實例（Instance）、單件（Item）等三大核心類別（Core classes）。
- 語彙集類型，除了直接採用BIBFRAME外，本研究也再用其它語彙集（諸如：DBpedia、Skos、Time、dcterms、rdf、schema、vra等），以及由本研究自行設計的語彙集（ASCDC），藉此擴展以完整描述古漢籍善本書目資訊。

類別與屬性	語彙集類型			總計（個）
	BIBFRAME	其他語彙集	ASCDC（本研究自行設計）	
類別	20	4	2	26
屬性	34	24	31	89

古漢籍善本知識本體的類別與屬性之語彙集類型

古漢籍善本知識本體

進行中研院傅斯年圖書館建置「古漢籍善本數位化資料庫國際合作建置計畫」內容轉置為LOD資料集所訂定

- 7,439 筆書目資源
- 140,668 筆資料三元組 (triples)
- 3 種核心類別 (Work, Instance, Item 依據 BIBFRAME 2.0)
- 自行建置或再利用12種外部語彙(ascdc, bibframe, dbpedia-owl, dc, dcterms, gndo, owl, rdf, rdfs, schema, skos, vra) 進行26種類別、89種屬性設計
- 鏈結再利用5種外部資源(AAT, DBpedia, LC, TGN, VIAF) · 達成資料加值功能

Work [29]

ascdc:classificationGrade, ascdc:collectionLevel, ascdc:dynasty, ascdc:FSNLibraryClassHeading, ascdc:FSNLibraryClassHeading1, ascdc:FSNLibraryClassHeading2, bf:agent, bf:arrangement, bf:contribution, bf:hasPart, bf:hasInstance, bf:hierarchicalLevel, bf:language, bf:mainTitle, bf:originDate, bf:relatedTo, bf:role, bf:subject, bf:title, dcterms:description, dcterms:identifier, dcterms:language, dcterms:issued, owl:sameAs, rdf:type, rdfs:label, rdfs:seeAlso, schema:image, schema:keywords

Instance [45]

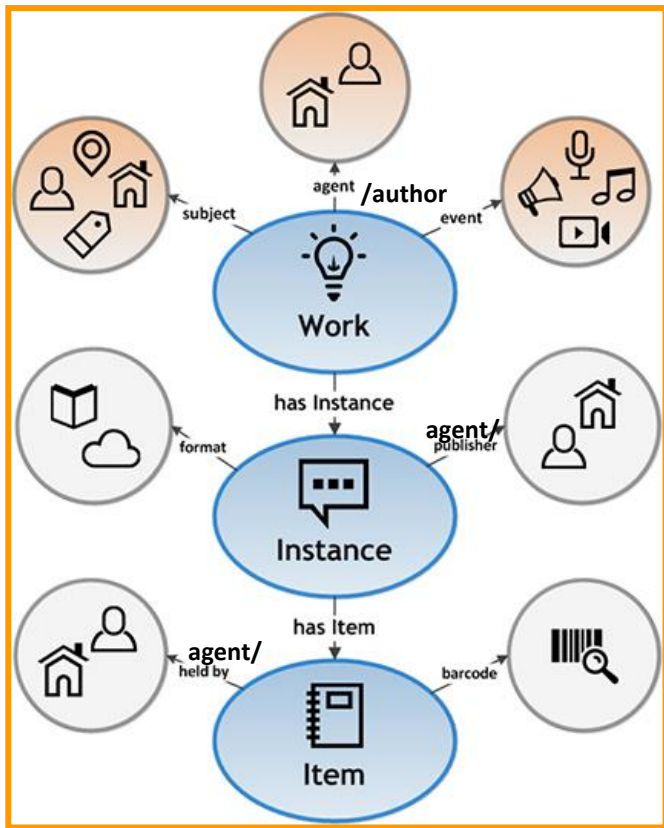
ascdc:accessoryType, ascdc:datePrinted, ascdc:dynasty, ascdc:erTi, ascdc:frameLineType, ascdc:illustrativeContent, ascdc:layoutSize, ascdc:linesPerPage, ascdc:moDing, ascdc:mouthOfBook, ascdc:place, ascdc:scriptType, ascdc:taboo, ascdc:tabooCharacters, ascdc:tabooPattern, ascdc:tabooType, ascdc:translationNote, ascdc:yuWei, bf:agent, bf:baseMaterial, bf:bookFormat, bf:colorContent, bf:contribution, bf:count, bf:date, bf:dimension, bf:editionStatement, bf:genreForm, bf:hasItem, bf:hasPart, bf:place, bf:relatedTo, bf:responsibilityStatement, bf:role, bf:title, bf:variantType, dcterms:description, dcterms:identifier, owl:versionInfo, owl:sameAs, rdf:type, rdfs:label, rdfs:seeAlso, schema:locationCreated, skos:note

Item [31]

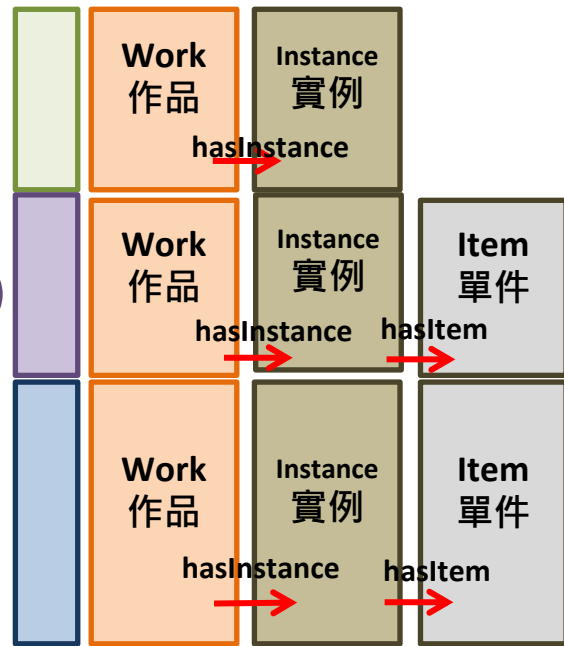
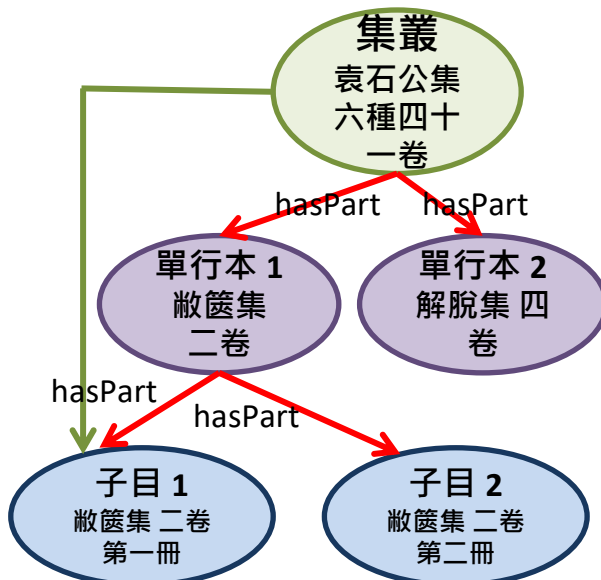
ascdc:acquisitionDate, ascdc:acquisitionTerms, ascdc:datePrinted, ascdc:projectNote, ascdc:seal, ascdc:sealCarvedType, ascdc:sealCharacters, ascdc:sealForm, bf:agent, bf:contribution, bf:custodialHistory, bf:date, bf:electronicLocator, bf:hasItem, bf:hasPart, bf:heldBy, bf:identifiedBy, bf:immediateAcquisition, bf:physicalLocation, bf:relatedTo, bf:role, dbpedia-owl:currentStatus, dbpedia-owl:owner, dc:source, dcterms:identifier, gndo:formerOwner, rdf:type, rdfs:label, schema:text, skos:note, vra:idCurrentAccession

古漢籍善本資料與BIBFRAME結構對應

BIBFRAME 2.0

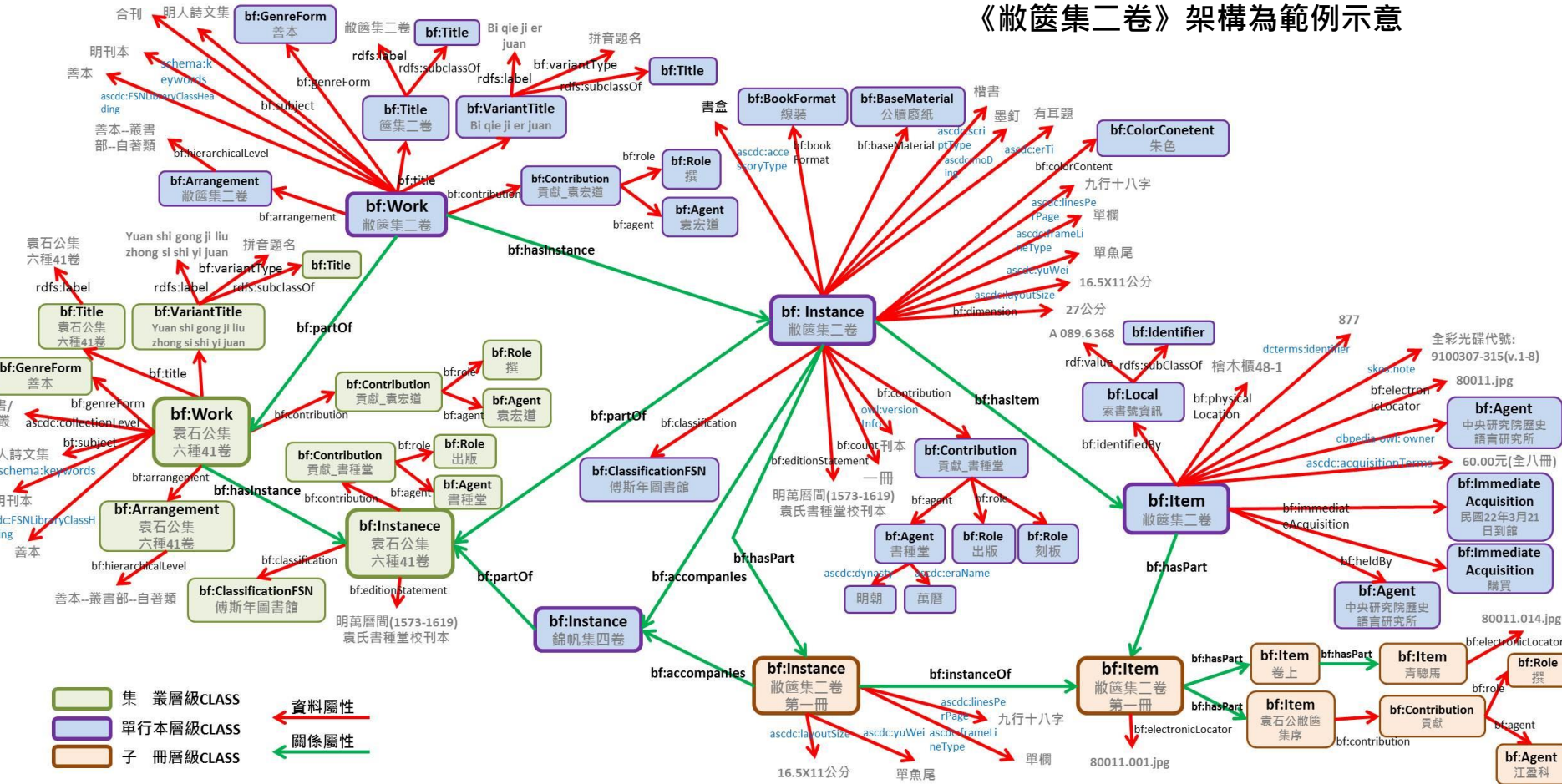


- 「古漢籍善本知識本體」(Chinese Rare Book Ontology · CRBO) 語意模型主要依據2016年由美國國會圖書館發布之「書目框架」(Bibliographic Framework · BIBFRAME) 2.0版本為基礎進行設計
- BIBFRAME模型架構不僅適用善本書目資訊描述，亦為目前即有書目資源本體中最为細膩、完整之資料模型設計之一



古漢籍善本知識本體的結構示意圖

以集叢《袁石公集六種四十一卷》中單行本《敞篋集二卷》架構為範例示意



古漢籍善本知識本體_類別/ Classes

Class	Label	Class in Bibframe
bf:Agent	代理者	Work, Instance, Item
bf:Arrangement	在地辨識號	Item
bf:BaseMaterial	書目實例	Instance
bf:BookForma	書目類型	Work
bf:ColorConte	顏色資訊	Instance
bf:Contributor	貢獻	Work, Instance, Item
bf:GenreForm	書目類型	Work
bf:Identifier	辨識碼	Item
bf:ImmediateAcquisition	取得資訊	Item
bf:Instance	書目實例	Instance
bf:Item	書目單件	Item
bf:Language	語言	Work
bf:Local	在地辨識號	Item

更細緻的屬性需求、滿足屬性的定義域與值域需求、便利資訊語意檢索的捷徑、特殊、在地化的書目資訊

Class	Label	Class in Bibframe
bf>Note	備註	Work, Instance, Item
bf:ParallelTitle	翻譯標題	Work
bf:Place	地點資訊	Instance
bf:Role	貢獻角色	Work, Instance, Item
bf:Title	標題	Work, Instance
bf:VariantTitle	異名標題	Work, Instance
bf:Work	書目作品	Work
ascdc:Seal	印記資訊	Item
ascdc:Taboo	避諱資訊	Instance
dbo:Image	圖像	Instance, Item
time:ProperInterval	時段實體	Instance
time:TemporalEntity	時間實體	Instance
skos:Concept	概念	Work, Instance, Item

註記

黑色類別: 採用自BIBFRAME模型

紅色類別: 再利用其他語彙或ASCDC自定之屬性

古漢籍善本知識本體_屬性 (節選 - Work層級)

Property	Label	Domain	Range	Value
bf:genreForm	類型	bf:Work	bf:GenreForm	善本
bf:subject	專題	bf:Work	N/A	明人詩文集
bf:arrangement	主題	bf:Work	bf:Arrangement	URI
bf:hierarchicalLevel	主題 (資訊)	bf:Arrangement	Literal	善本--叢書部--自著類
bf:subject	層級	bf:Work	N/A	合刊圖書
schema:keywords	關鍵詞	bf:Work	Literal	明刊本
bf>Title	主要題名	bf:Work	bf>Title	URI
rdfs:label	主要題名 (名稱)	bf>Title	Literal	袁石公集六種41卷
bf>Title	拼音題名	bf:Work	bf:VariantTitle	URI
rdfs:label	拼音題名 (名稱)	bf:VariantTitle	Literal	Yuan shi gong ji liu zhong si shi
bf:variantType	拼音題名 (類型)	bf:VariantTitle	Literal	拼音題名
rdfs:subclassOf	拼音題名 (類型)	bf:VariantTitle	bf>Title	Title

Property	Label	Domain	Range	Value
skos:note	簡述	bf:Work	Literal	袁宏道著作全集
bf:contribution	貢獻	bf:Work	bf:Contribution	URI
bf:agent	貢獻(人名)	bf:Contribution	bf:Agent	袁宏道
bf:role	貢獻(角色)	bf:Contribution	bf:Role	撰
ascdc:dynasty	朝代	bf:Agent	time:TemporalEntity	明朝
ascdc:eraName	年號	bf:Agent	time:TemporalEntity	萬曆
bf:language	語言	bf:Work	bf:Language	中文
dcterms:language	語言代碼	bf:Work	Literal/ISO639	chi
dcterms:identifier	ID號碼	bf:Work	String	20562
bf:hasInstance	所屬Instance	bf:Work	bf:Instance	袁石公集六種41卷(Instance)
bf:hasPart	所屬單行本	bf:Work	bf:Work	敝篋集二卷(Work)

黑色類別: 採用自BIBFRAME模型

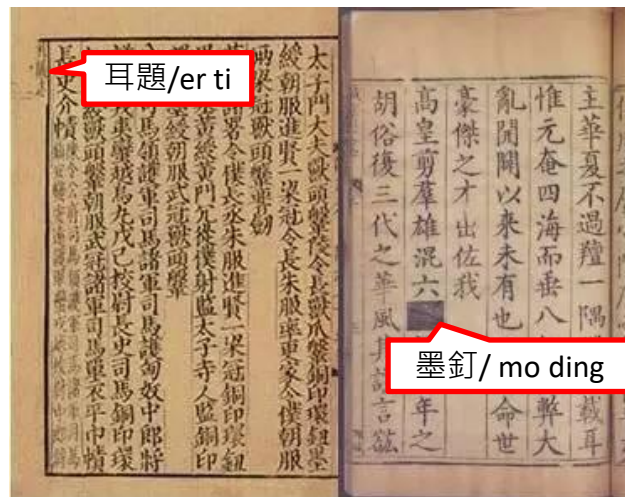
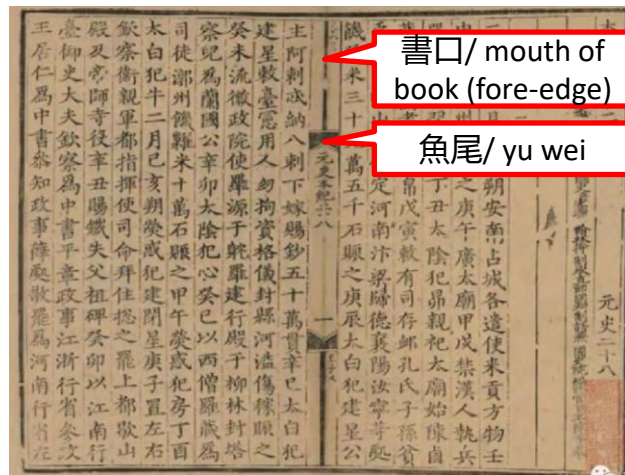
紅色類別: 再利用其他語彙或ASCDC自定之屬性

古漢籍善本知識本體_屬性 (節選 - Instance層級)

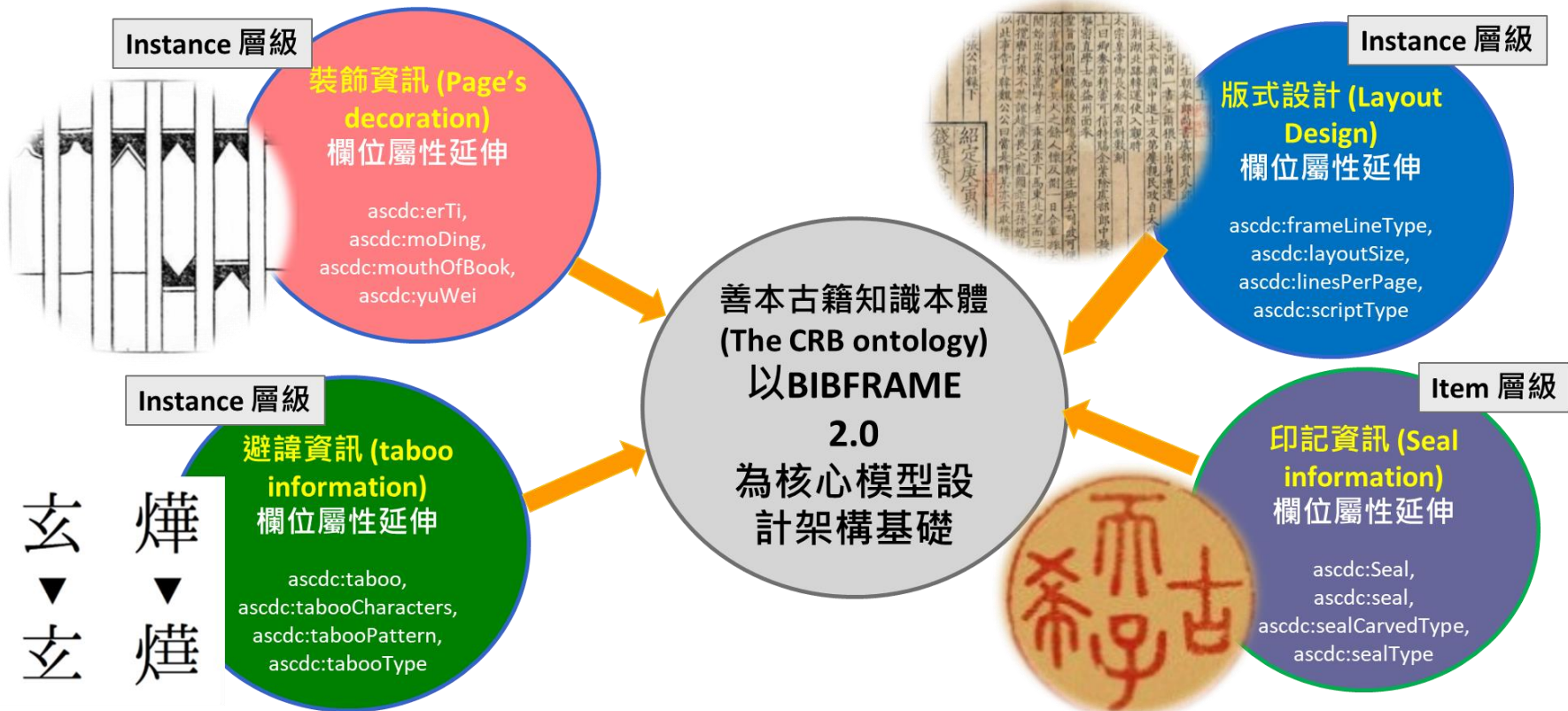
Property	Label	Domain	Range	Value
ascdc:linesPerPage	行格	bf:Instance	Literal	九行十八字
bf:dimensions	外型高廣	bf:Instance	Literal	27x14公分
ascdc:layoutSize	版框高廣	bf:Instance	Literal	16.5x11公分
ascdc:frameLineType	邊欄	bf:Instance	Literal	單欄
ascdc:mouthOfBook	書口	bf:Instance	Literal	白口
ascdc:yuWei	魚尾	bf:Instance	Literal	單魚尾
ascdc:moDing	墨釘	bf:Instance	Literal	墨釘
ascdc:erTi	耳題	bf:Instance	Literal	有耳題
bf:bookFormat	裝訂	bf:Instance	bf:BookFormat	蝴蝶裝
ascdc:accessoryType	裝潢	bf:Instance	Literal	書盒
bf:baseMaterial	材質	bf:Instance	bf:BaseMaterial	公牘廢紙
bf:colorContent	墨色	bf:Instance	bf:ColorContent	朱色
ascdc:scriptType	書體	bf:Instance	skos:Concept	楷書
ascdc:illustrativeContent	圖像	bf:Instance	skos:Concept	圖
ascdc:translationNote	譯文	bf:Instance	Literal	藏文譯中文

黑色類別: 採用自BIBFRAME模型

紅色類別: 再利用其他語彙或ASDC自定之屬性



古漢籍善本知識本體：BIBFRAME的擴充_屬性



「古漢籍善本知識本體」在屬性元素延伸設計上，主要針對「實例」層級中的「裝飾」(page's decoration)、「版式設計」(layout design)、「避諱」(taboo)及「單件」層級中的「印記」(seal)等相關欄位資訊進行新增屬性設計

古漢籍善本資料集的資料與語意加值

- 鏈結內部或外部已發布之異質資料(如控制詞彙、索引典資源等)，達成資料加值功能(data enrichment)
- 藉由將原有文字資訊轉換為具URI標示之物件(thing)資源，使其可被檢索並鏈結其他外部資源
- 經由資料加值步驟，激發、創造新知識及拓展研究、詮釋新視角。豐富、增加原有資料內容，並達成資料間以語意方式串連等目的。

Getty Vocabularies: LOD
standard script (Chinese script)
Source: <http://vocab.getty.edu/aat/300343590>

Subject (81) Predicate Object All

Website | Hierarchy | Download in:

Inference: Explicit only

Statements in which the resource exists as a subject.

Predicate Object

rdf:type gvp:Concept

AAT藝術與建築索引典

書體
ascdc:scriptType

A Digital Library of Chinese Rare Books

古漢籍善本
數位化資料庫
檢索系統

http://data.ascdc.tw/Book/crb/Book_Work/114487 (嘉興府志)

出版地
schema:locationCreated

等同資源
owl:sameAs

版權
bf:heldBy

VIAF 虛擬國際權威檔
Virtual International Authority File

中央研究院歷史語言研究所
中央研究院 歷史語言研究所
Institute of history and philology Taipei
Academia Sinica (Taipei) Institute of History and Philology
Institute of History and Philology of Academia Sinica organization
VIAF ID: 124359848 (Corporate)
Permalink: <http://viaf.org/viaf/124359848>

LIBRARY LIBRARY OF CONGRESS

Everything

The Library of Congress > Linked Data Service > BIBFRAME Works

Bibframe Work

Title
jia xing fu zhi
嘉興府志(浙江)

Type
Text

Contribution

LOC書目詞彙

Getty Vocabularies: LOD

Zhejiang

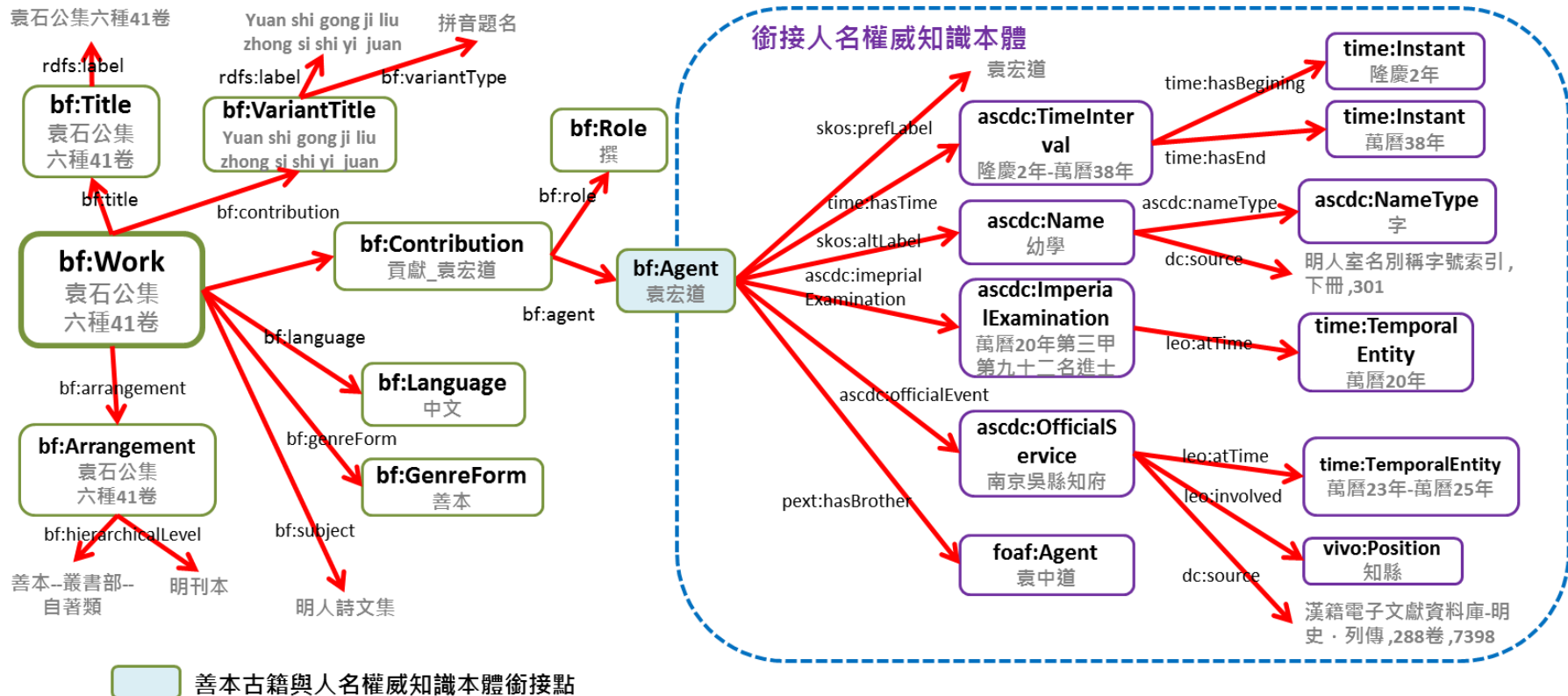
Source:<http://vocab.getty.edu/tgn/7002076>

Subject (69) Predicate Object All

TGN

古漢籍善本語意加值- 中研院內部資源 (資料集) 鏈結

再利用古漢籍善本資源中創作者姓名資源鏈結ASCDC明清人名權威檔LOD資料集，達成跨資料集內容加值功能



古漢籍善本資料增值- 外部資源鏈結

藉由設計、創建自訂之資料集URIs，以鏈結網路外部即有控制詞彙/資源，由此分享鏈結資源實體內容

鏈結美國國會圖書館BIBFRAME資料集 (包含BIBFRAME Work、BIBFRAME Instance等層級) 達成跨資料及查詢



古漢籍善本書目鏈結開放資料集與數位人文研究

以問題意識為基礎的資料檢索、分析及視覺化

	檢索範例	研究議題	研究階段
1	善本資料集中收錄的各單行本的出版年份，分別位於哪些皇帝統治期間？各有多少數量？	目錄學-出版史-思想文化	研究初步查考階段，輔助研究者從大方向掌握各時期書籍出版數量分布
2	善本資料集中收錄的各單行本的出版地點，分別位於今日的那些地區？各有多少數量？	目錄學-出版史-書籍史	研究初步查考階段，藉由統計結果呈現作為研究主題
3	<ul style="list-style-type: none">• 善本古籍資料集中有那些方志類書籍、地域上之分布及其四部主題分類？• 同一地名，不同朝代，呈現怎樣的名目變化？其分類體例又何不同？	目錄學-出版史-學術思想史 古典文獻、歷史學範疇	研究拓展的階段
4	古漢籍善本LOD資料集中，各單行本分別有哪些貢獻者？又從曾共同參與同一單行本的次數來看(含撰述、序跋題記...)，貢獻人物彼此間關聯性如何？	目錄學-出版史-合作關係、 文人史-文人社群-學術思潮、善本書目著作群體社會關係研究 (SNA)	完整研究問題，並具清晰問題意識
5	古漢籍善本資料集中所收錄的“嘉興府志”版本為何，又該作品在美國國會圖書館典藏目錄(Library of Congress)中還有那些其他版本？	善本書目版本研究、 跨資料集查詢	基礎問題，輔助文獻學學者進行資料查考，具協助研究的價值
6	善本古籍中那些子目書籍版式行格為十行二十字？這類書籍所屬之分類類目及出版朝代為何？	善本版式裝幀研究	深度研究問題拓展，輔助進行古籍出版景氣環境、出版經費、出版策略歷史演變

古漢籍善本資料增值- 外部資源鏈結

檢索範例 (No. 5)

古漢籍善本資料集中所收錄的“嘉興府志”版本為何，又該作品在美國國會圖書館典藏目錄(Library of Congress)中還有那些其他版本？

研究議題

善本書目版本研究、
跨資料集查詢

研究階段

基礎問題，輔助文獻學學者
進行資料查考，具協助研究的
價值

Triples

RESULTS ▾

crbWork	crbLabel	crbEdition	locWork	locEdition
http://data.ascdc.tw/Book/crb/Book_Work/114487	嘉興府志十六卷	清康熙六十年(1721)刊本	http://id.loc.gov/resources/works/19346376	[民國傳抄明嘉靖二十八年本]
http://data.ascdc.tw/Book/crb/Book_Work/114487	嘉興府志十六卷	清康熙六十年(1721)刊本	http://id.loc.gov/resources/works/19346374	[清康熙六十年刻本]
http://data.ascdc.tw/Book/crb/Book_Work/114487	嘉興府志十六卷	清康熙六十年(1721)刊本	http://id.loc.gov/resources/works/19709572	[清光緒四年刻本]
http://data.ascdc.tw/Book/crb/Book_Work/114487	嘉興府志十六卷	清康熙六十年(1721)刊本	http://id.loc.gov/res	鏈結至LOC BIBFRAME資料集
http://data.ascdc.tw/Book/crb/Book_Work/114487	嘉興府志十六卷	清康熙六十年(1721)刊本	http://id.loc.gov/resources/works/19709571	[清道光二十年刻本]
http://data.ascdc.tw/Book/crb/Book_Work/114487	嘉興府志十六卷	清康熙六十年(1721)刊本	http://id.loc.gov/resources/works/19346375	[清同治六年刻本]
http://data.ascdc.tw/Book/crb/Book_Work/114487	嘉興府志十六卷	清康熙六十年(1721)刊本	http://id.loc.gov/resources/works/19730440	[明萬曆年間刻本]

資料調和外部資源範例

藉由鏈結美國國會圖書館BIBFRAME資料集等外部資援，使用者可檢索出與「古漢籍善本資料集」中某特定書目資源相關之其他書目版本，援此達成原始「古漢籍善本資料集」內容增值工作。

建立SPARQL endpoint檢索端點並創建語意檢索範例模組，達成機器及人為資料檢索可行性

SPARQL語意檢索結果資料視覺化呈現模組

檢索範例 (No. 2)

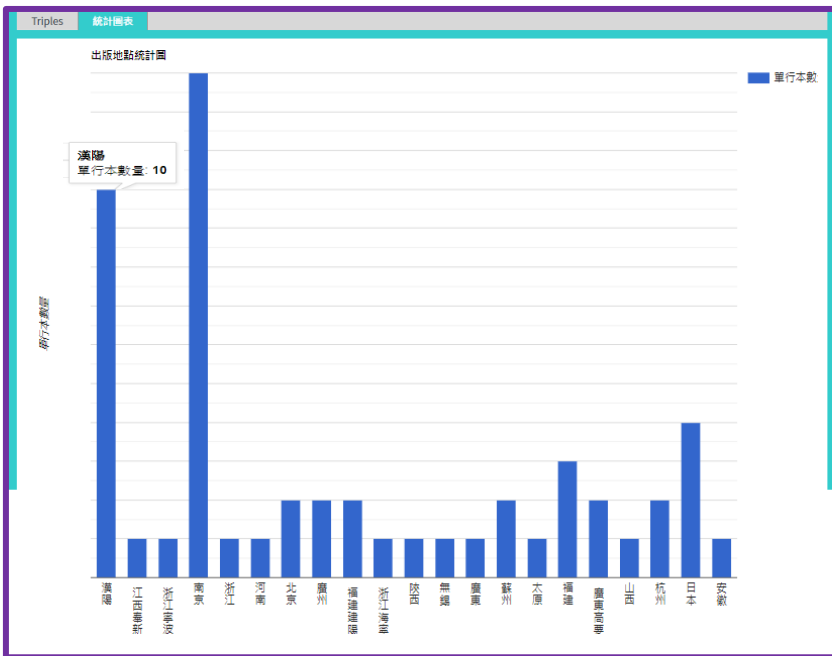
善本資料集中收錄的各單行本的出版地點，分別位於今日的那些地區？各有多少數量？

研究議題

目錄學-出版史-書籍史

研究階段

研究初步查考階段，藉由統計結果呈現作為研究主題



	Subject	Property	Object
善本古籍書目_URI	?	ascdc:place (出版地點)	?
單行本_Instance_出版地點	FILTER (Contains(STR(?s), 'Instance')) count(?s) as ?countO 寫入過濾指令，檢索Instance層級中有資訊之部分。 其次將檢索之結果，以count指令進行統計，以利統計結果於countO檢索欄中呈現		

檢索後經統計地名及數字，再以視覺化方式呈現結果

在研究意義上，藉由「出版地點」檢索，從其地名即可分類、歸納書目版本學研究中慣常使用的「地點版本」資訊，如：建陽版、漢陽版等，藉此展現各地刊刻事業崛起的先後順序。

SPARQL語意檢索結果資料視覺化呈現模組

檢索範例 (No. 3)

善本古籍資料集中有那些方志類書籍、地域上之分布及其四部主題分類？
可以再延伸探究：同一地名，不同朝代，呈現怎樣的名目變化？其分類體例
又何不同？

研究議題

目錄學-目錄分類
古典文獻、歷史學範疇

研究階段

研究初步查考階段，
藉由統計結果呈現作
為研究主題



	Subject	Property	Object
單行本_ Work_ 主題	?	bf:subject	?
單行本_ Work_ 四部主題	(單行本作品 層級資源實例)	ascdc:FSNLibraryClassHeading	?
單行本_ Work_ 題名	(單行本作品 層級資源實例)	skos:prefLabel	?
單行本_ Work_ 主題	FILTER (str(?subject) = '方志') 寫入過濾指令，檢索Work層級中，主題值為“方志”者		

「方志」屬於「地理類」，「地理類」在古典文獻被歸類在「史部」。同一地，不同朝代，會因行政區的變動、行政位階的不同，產生方志書寫時體例的變化，從書名就能觀察得出變動的狀況

古漢籍善本書目鏈結開放資料集與數位人文研究

檢索範例 (No. 4)

古漢籍善本LOD資料集中，各單行本分別有哪些貢獻者？又從曾共同參與同一單行本的次數來看(含撰述、序跋題記...)，貢獻人物彼此間關聯性如何？

研究議題

- 目錄學-出版史-合作關係
- 文人史-文人社群-學術思潮、善本書目著作群體社會關係研究

研究階段

完整研究問題，並具清晰問題意識。以建構文人網絡，探索文人社群的文學研究方式為主。

中央研究院數位文化中心
鏈結開放資料平台

資料查詢 SPARQL查詢 資料集下載

查詢範例，可點選參考

範例8. 上述作品在古漢籍善本LOD資料集中，所收錄的版本為何？

範例9. 上述作品出版後的其他相關貢獻者有誰？著作方式為何？在人名權威LOD資料集、古漢籍善本LOD資料集中的URI分別為何？

範例10. 在古漢籍善本LOD資料集中，各單行本分別有哪些貢獻者？又，從曾共同參與同一單行本的次數來看(包含撰述、序跋題記……)，貢獻人物彼此間的關聯性如何？

範例11. 善本古籍資料集中有哪些方志類書籍及其主題分類？又這些書籍在美國國會圖書館典藏目錄(Library of Congress)中的書籍分類資訊為何？

範例12. 善本古籍資料集中所收錄的“嘉興府志”版本為何，又該作品在美國國會圖書館典藏目錄(Library of Congress)中還有那些其他版本？

範例13. 善本古籍中那些子目書籍版式行格為十行二十字？這類書籍所屬之分類類目及出版朝代為何？

範例14. 善本古籍中那些字行古高厚寬行何時早有何十字？這類書籍所屬之分類類目及出版朝代為何？

範例14. 善本古籍中那些單行本子目書籍同時具有雙魚尾、雙欄等版式裝飾？這類書籍所屬出版朝代資訊為何？

SPARQL QUERY 語意輸入框

```
1 PREFIX bf: <http://id.loc.gov/ontologies/bibframe/>
2 PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
3 PREFIX dcterm: <http://purl.org/dc/terms/>
4 SELECT ?uri ?contributor ?role ?label
```

SPARQL範例查詢說明

- 1) 篩選出所有S為單行本work與InstanceURI，P為bf:contribution、bf:agent的O值。轉出這些O值對應的literal人名。
- 2) 承上，篩選出上述有值的單行本work與InstanceURI為S，P為dcterms:identifier的O值。

SPARQL語意檢索結果行列顯示

uri	contributor	role	label
http://data.ascd.edu.tw/Book/crb/Book_Work/105118	施世倫	撰	南堂詩鈔十二卷南堂詞賦一卷
http://data.ascd.edu.tw/Book/crb/Book_Work/100881	如松	撰	竹窠雅譚四卷
http://data.ascd.edu.tw/Book/crb/Book_Work/112866	張岳	撰	小山類稿選二十卷
http://data.ascd.edu.tw/Book/crb/Book_Work/105106	汪沆	著	槐塘詩文稿二十卷
http://data.ascd.edu.tw/Book/crb/Book_Work/112026	馮敏昌	修	孟縣志十卷
http://data.ascd.edu.tw/Book/crb/Book_Work/101000	趙廷瑞	編	南齋會景編十卷
http://data.ascd.edu.tw/Book/crb/Book_Work/102013	顧溥	著	萍留草一卷續草一卷
http://data.ascd.edu.tw/Book/crb/Book_Work/102266	謝定夫	撰	慈溪謝定夫詩話十卷

- 本項研究命題嘗試找出善本書目創作者間人際互動關係，藉此揭示中國傳統文人間在書籍創作中的群聚性或離散性。

- 以SPARQL語意檢索方式，查詢「古漢籍善本鏈結開放資料」中集單行本貢獻者、貢獻方式、書目資訊，總計獲得1,469位，作為後續進行SNA網路分析之基礎。

古漢籍善本書目鏈結開放資料集與數位人文研究

檢索範例 (No. 5)

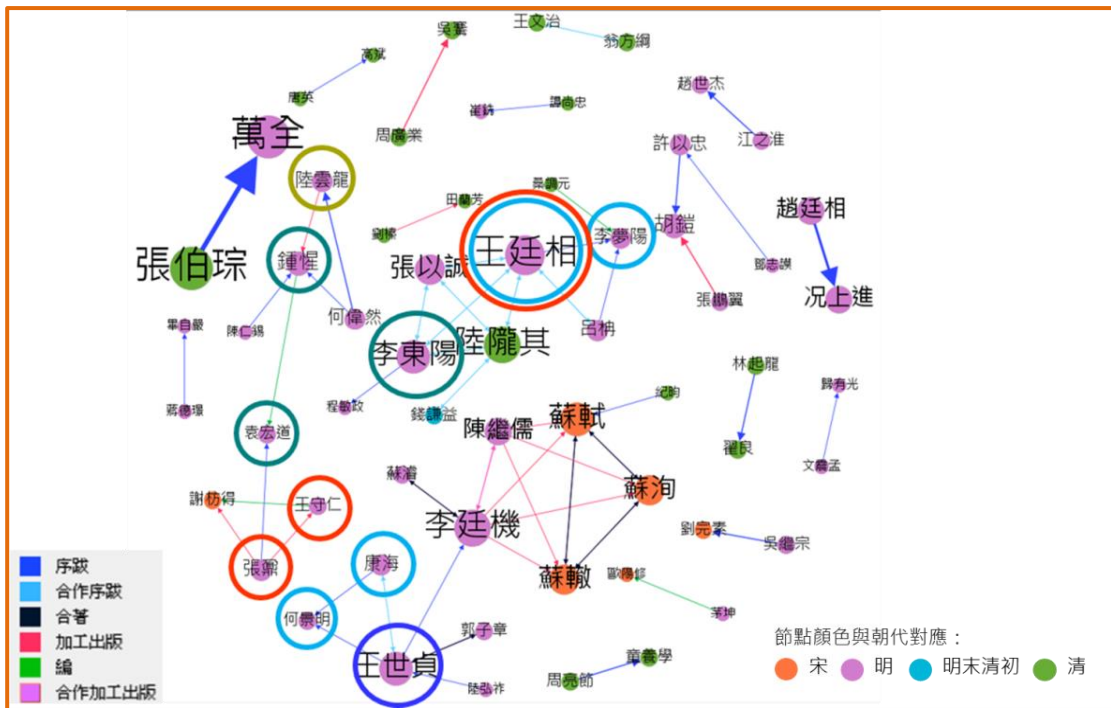
古漢籍善本LOD資料集中，各單行本分別有哪些貢獻者？又從曾共同參與同一單行本的次數來看(含撰述、序跋題記...)，貢獻人物彼此間關聯性如何？

研究議題

- 目錄學-出版史--合作關係
- 文人史-文人社群-學術思潮、善本書目著作群體社會關係研究

研究階段

完整研究問題，並具清晰問題意識。以建構文人網絡，探索文人社群的文學研究方式為主。



- 利用SPARQL語意檢索結果形成資料集，再運用Gephi分析工具，顯示善本書目共同創作、貢獻參與人數、創作類型。
- 由脈絡圖顯示：從宋代三蘇連結出陳繼儒、李廷機，復連結至王世貞、何景明、康海，有別於歷來「文人集團」（如前七子/後七子）的劃分與理解。由此結果可跳脫既有文人集團歸類印象，重新注意文人具體真實的往來關係，為文人史、文學社會學之研究，提供可能發展之新面向。

結語

- 本研究提供更開放、更結構化的資料集，突破傳統書目資料孤島(silos)的問題。 [[Linked Open Data](#)]
- 基於BIBFRAME 2.0為基礎，為古漢籍善本書目資料(1)提出「類別」與「屬性」擴展，完善領域知識；(2)經由語意擴充(semantic enrichment)方式建構資料增值，讓鏈結開放資料雲(The Linked Open Data Cloud)的資料集彼此間具有語意鏈結。
- 基於上述的擴充，可以整合其他資料集的內容，也可以讓本身資料集讓他人再利用。
- 示範以問題意識為基礎之資料檢索、分析及視覺化，提供人文研究者依不同研究階段、不同研究主題，更有彈性的語意查詢方法。並以這些資料的分析結果，進行數位人文研究。 [[SPARQL 語意查詢樣版](#)]

致謝

- 林玟君(中研院數位文化中心博士後研究員；
書志學研究員)
- 盧履彥(中研院數位文化中心鏈結開放資料
實驗室成員)

中央研究院數位文化中心 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



THANK YOU!

