

# 面向循证实践的中文古籍 数据模型设计与实现

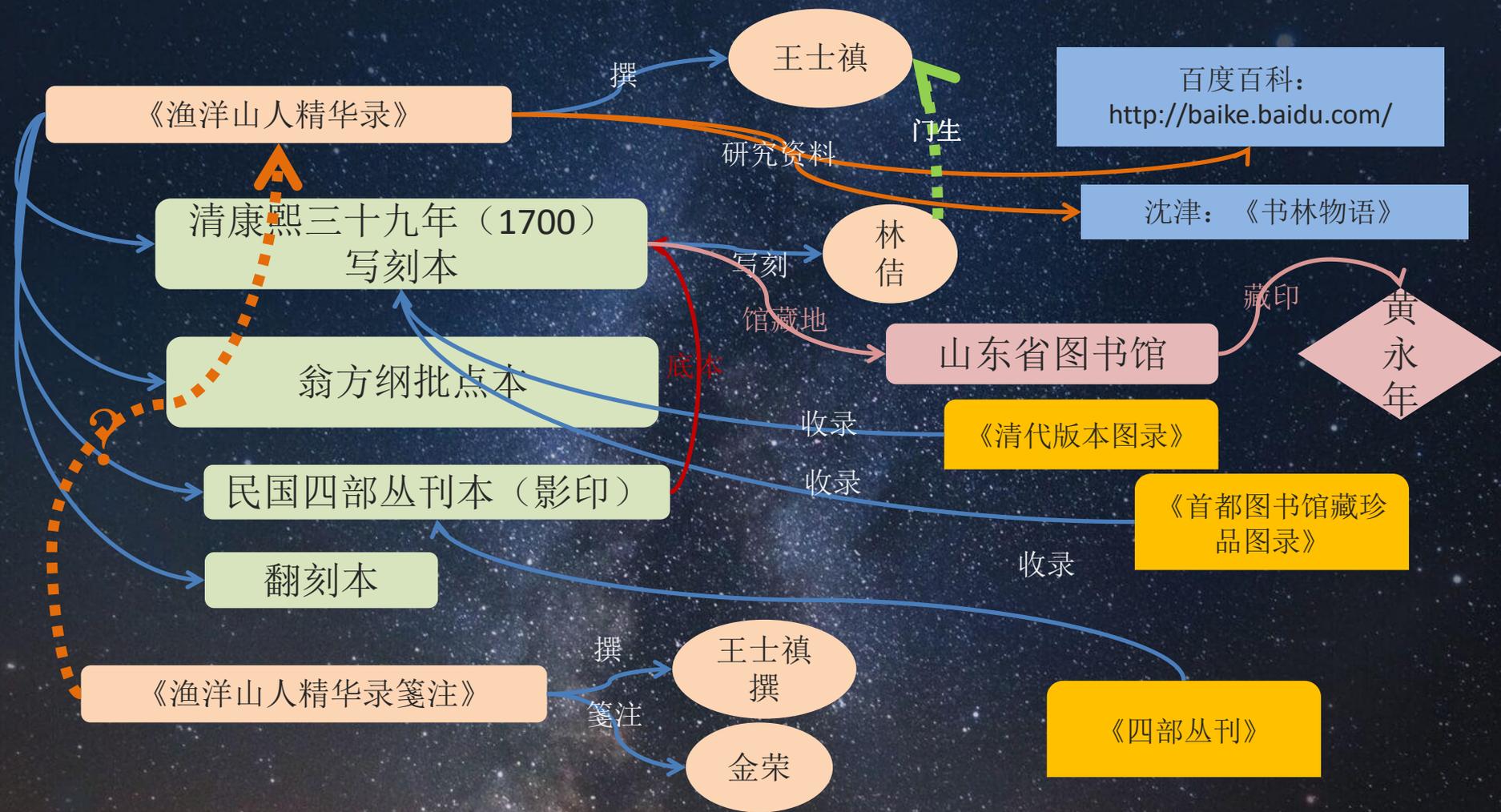
上海图书馆高级工程师  
数字人文项目主管 夏翠娟  
2018年6月@\*\*\*\*大学

# 1. 什么是古籍循证



# 定义

以古籍目录的记载和古籍文献中的内容作为依据，将科学的研究方法与研究人员的经验结合起来，解决特定研究问题的过程，可称之为“古籍循证”。



# 源流探析

## 韩诗外传

是一部记述中国古代史实、传闻的著作。共十卷。由360条轶事、道德说教、伦理规范以及实际忠告等不同内容的杂编，一般每条都以一句恰当的《诗经》引文作结论，以支持政事或论辩中的观点，就其书与《诗经》联系的程度而论，它对《诗经》既不是注释，也不是阐发。《韩诗外传》以儒家为本，因循损益、以传资政，从礼乐教化、道德伦理等方面阐发了其思想。

1123-1202: 洪迈《容斋随笔》

1599-1659: 毛晋《津逮秘书》

1034: 张观、李淑、宋祁等《崇文总目》

1599-1659: 毛晋《亦有生斋》

1733: 《四库全书总目提要》

1747-1823: 赵怀玉《亦有生斋》

1811-1871: 莫友芝《邵亭知见传本书目》

1875: 吴棠《畿辅丛书》

1935-1937: 王云五《丛书集成》

# 构建古籍循证的证据链

Constriction of the evidence chains for research

- 物理证据 (Physical Evidence)
- 历史证据 (Historical Evidence)
- 内容证据 (Content Evidence)
- 关联证据 (Relational Evidence)

夏翠娟, 林海青, 刘炜. 面向循证实践的中文古籍数据模型研究与设计. 中国图书馆学报, 2017 (6)

# 传统的古籍循证方法

搜集资料-->阅读资料-->比较遴选-->分门  
别类-->深度阅读-->提炼观点-->发现未知--  
>查找更多资料... ..

# 机器智能时代的古籍循证方法

搜集资料：大数据、长数据、小数据

自动化、大规模、可共享

分析统计：聚类、比较、统计、分析

自动化、可视化、全景式

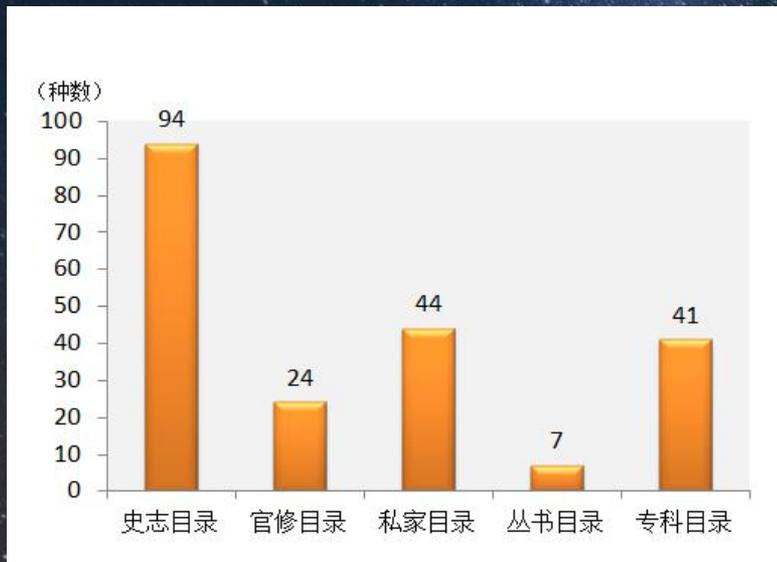
## 2. 古籍循证的需求

# 领域应用

- 书目控制
- 版本学
- 分类学
- 校勘学
- 人文社会研究领域

# 古籍目录

# 古籍数据库



中華經典古籍庫



全國古籍普查登記基本數據庫

全國漢籍データベース

日本所藏中文古籍數據庫

# 古籍联合目录及循证平台

- 不仅是现存古籍的联合目录，还是已散佚古籍的互联网发布平台
- 不仅是文献库，还是数据库
- 不仅是数据库，还是知识库
- 不仅是古籍目录，还是古籍数字全文库
- 不仅保存已出版的知识，还保存用户头脑中的知识

# 3. 数据模型设计



# 目标

- 容纳现存的古籍知识的同时，也要容纳已经散佚的古籍知识，还要容纳专家的研究成果
- 实现古籍的书目控制，深度揭示古籍不同版本之间的关系
- 深度揭示作者、批校题跋者、刻工、收藏者的生平、作品、著述，以及相互之间的关系
- 支持版本知识、藏印知识、避讳字知识的结构化、语义化

# 古代目录内容结构分析

表1 古代目录的内容结构分析

题名	分类	小序	题名	著者	其他责任者	解题	辑录
汉书·艺文志	有	有	有	有			
隋书·经籍志	有	有	有	有	有		
旧唐书·经籍志	有		有	有	有		
新唐书·艺文志	有		有	有	有		
宋史·艺文志	有		有	有	有		
明史·艺文志	有		有	有	有		
清史稿·艺文志	有		有	有	有		
崇文总目	有	有	有	有	有	有	
文献通考·经籍考	有		有	有	有	有	有
四库全书总目	有	有	有	有	有	有	
郡斋读书志	有	有	有	有	有	有	
直斋书录解题	有		有	有	有	有	
读书敏求记	有		有	有	有	有	
邵亭知见传本书目	有		有	有	有	有	
韶宋楼藏书志	有		有	有	有	有	有
铁琴铜剑楼藏书目录	有		有	有	有	有	
士礼居藏书题跋记	有		有	有	有	有	有
百宋一廬书录	有		有	有	有	有	
贩书偶记	有		有	有	有	有	

# 现代联合目录内容结构分析

表 2 当代主要古籍联合目录

题名	出版时间	藏书机构	收录范围	数量
中国丛书综录	1959-1962 初版, 中华书局; 1982 再版, 上海古籍出版社	41 家		丛书 2_797 种, 子目 7 万多条
中国地方志联合目录	1985 年, 中华书局	190 家	自南朝宋至 1949 年的方志	8_200 余种
中国古籍善本书目	1989 年 10 月至 1998 年 3 月, 上海古籍出版社	781 家	清乾隆前以及至辛亥革命前有特殊价值的刻本、抄本、稿本、校本	约 6 万多种
中国古籍总目	2009 年, 中华书局、上海古籍出版社	1_182 家		约 20 万种
中国家谱总目	2009 年, 上海古籍出版社	597 家 608 个姓氏	至 2003 年的家谱	52_401 种

# 各国国家图书馆 关联书目数据及其数据模型





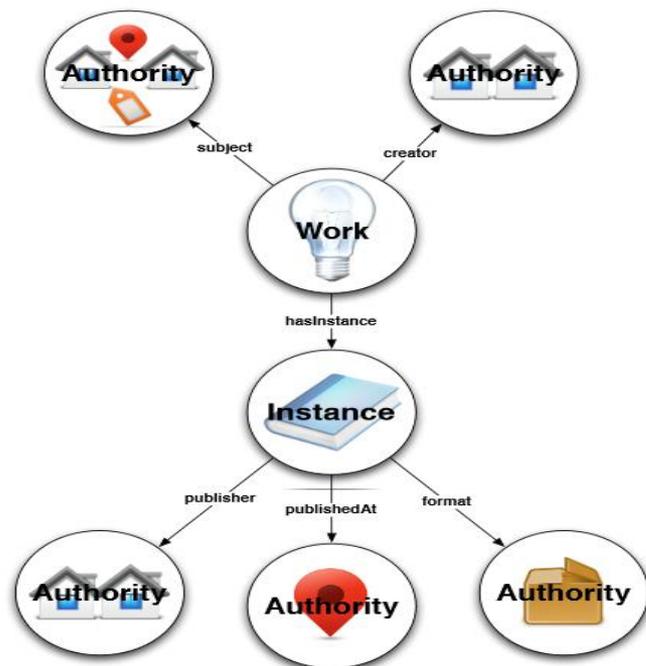
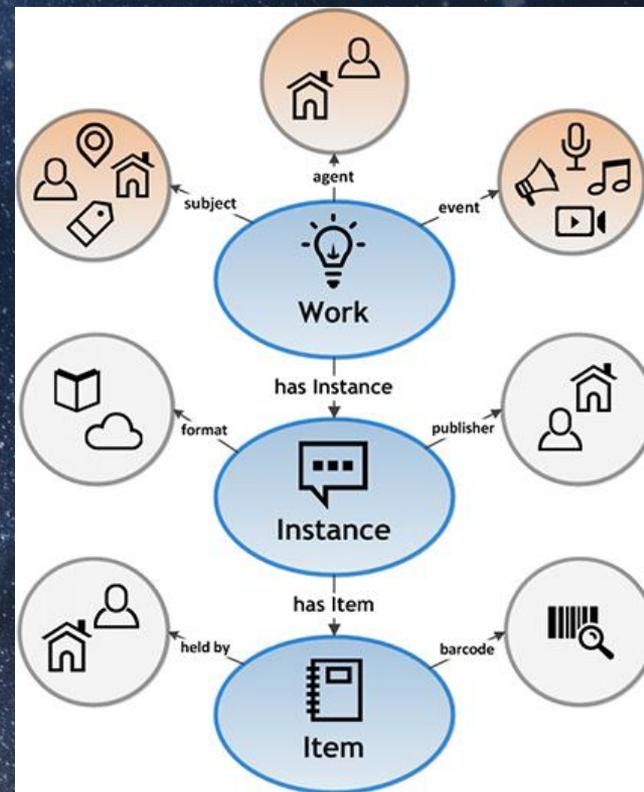


Figure 1: A graphical representation of the BIBFRAME Linked Data model defining the relation between Work and Instance resources and their contextualization to Web addressable Authority resources.

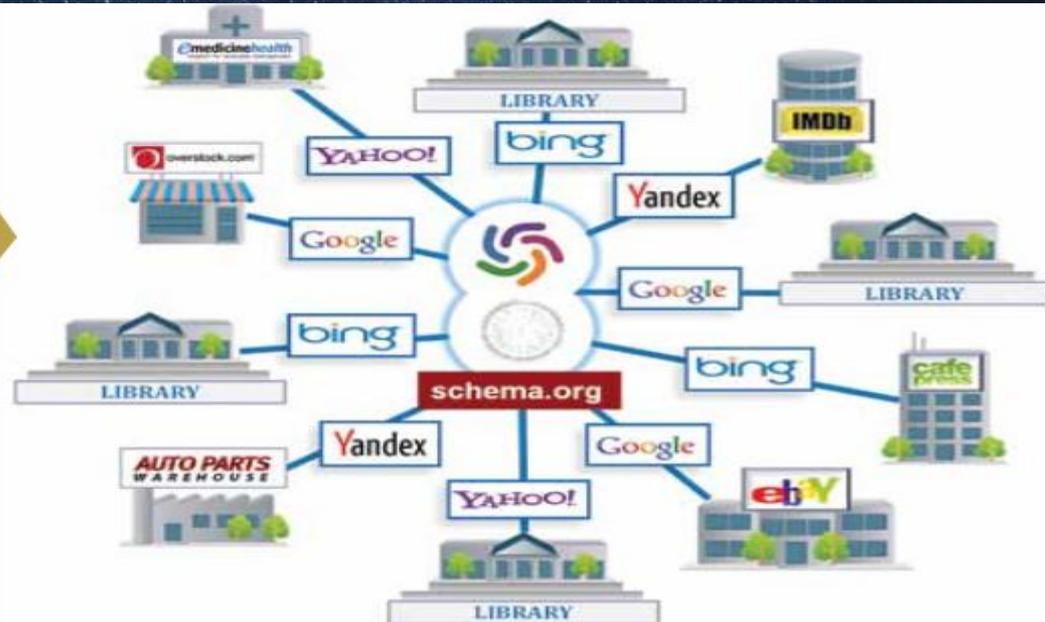


## 美国国会图书馆 (Library Of Congress) 书目框架 (BIBFRAME)

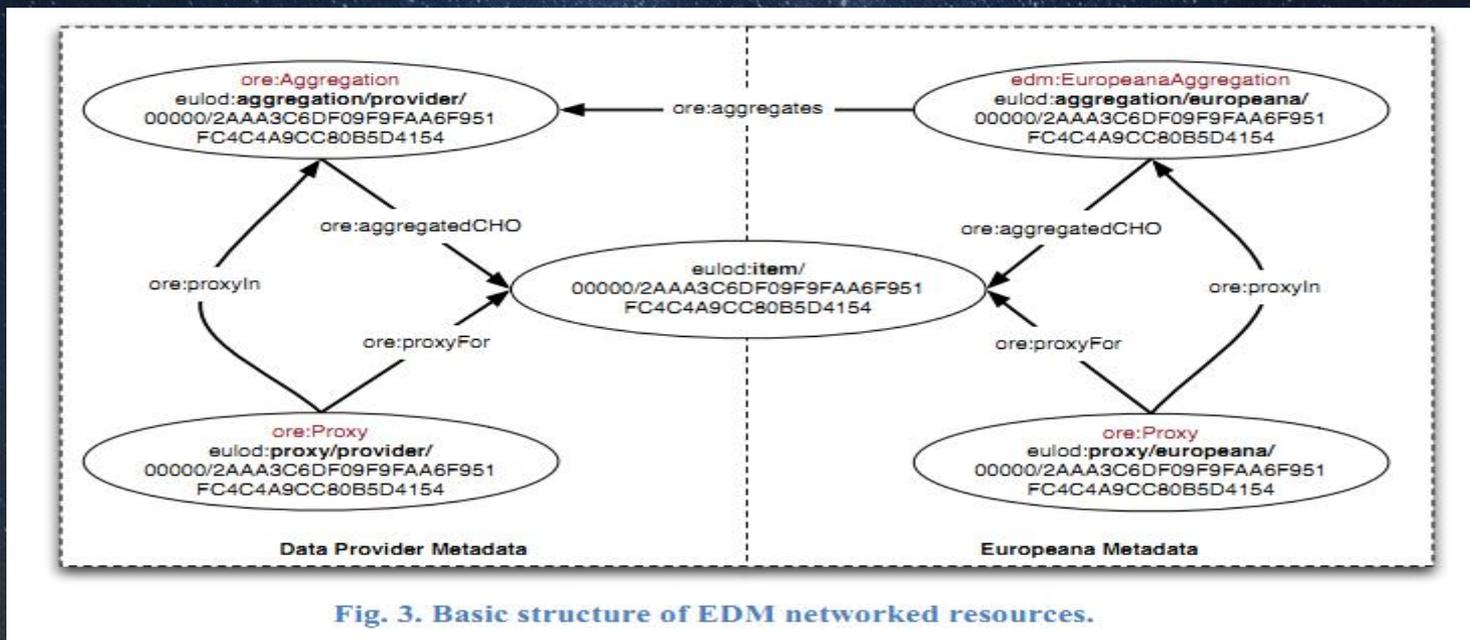
<http://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf> <https://www.loc.gov/bibframe/docs/bibframe2-model.html>

## WORLDCAT.ORG & SCHEMA.ORG

Using linked data and the Schema.org vocabularies as a starting point, rich domains like libraries, retailers, publishers, governments, and scientists can surface in this webscale interpretation with more context and clearer intent.



OCLC  
<http://www.oclc.org/data.en.html>



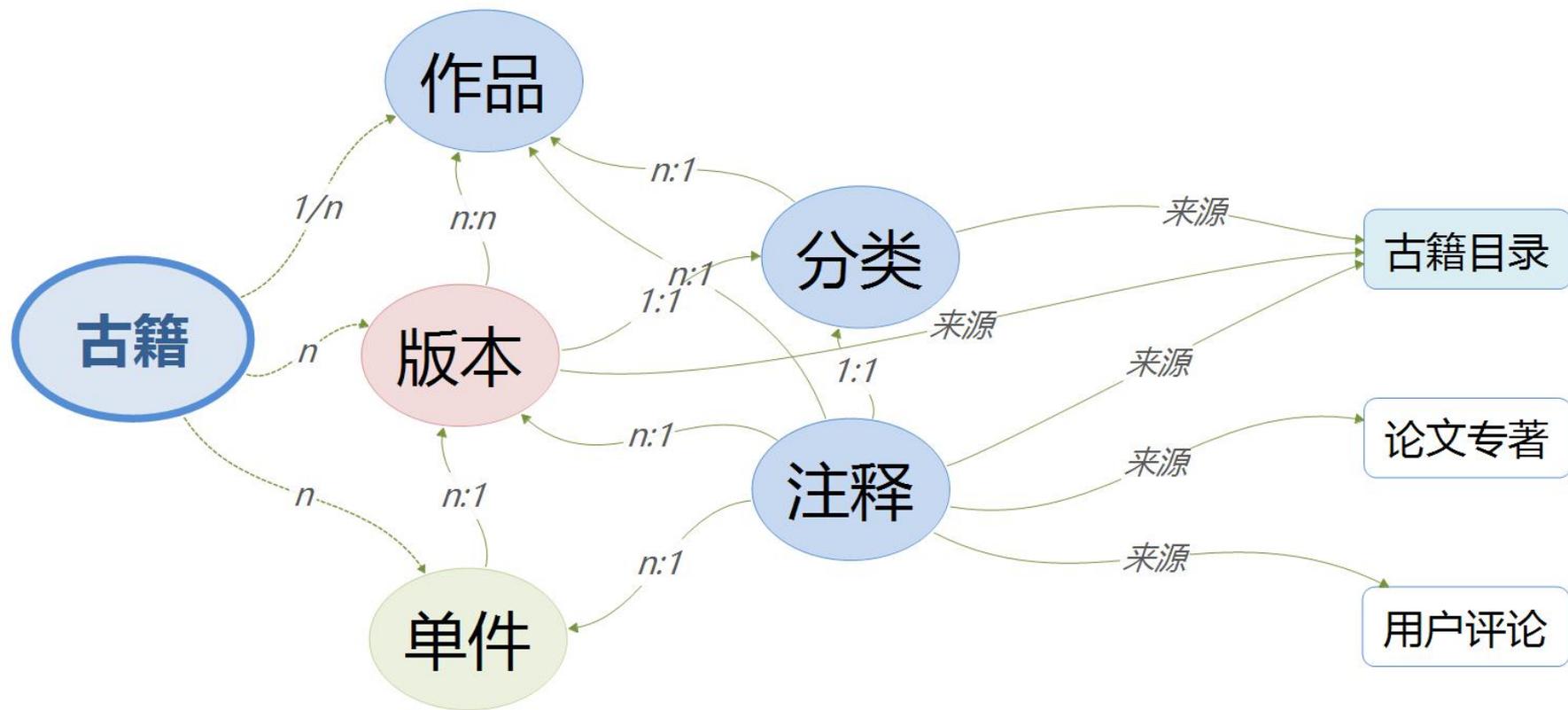
## 欧洲数字图书馆 (Europeana) EDM

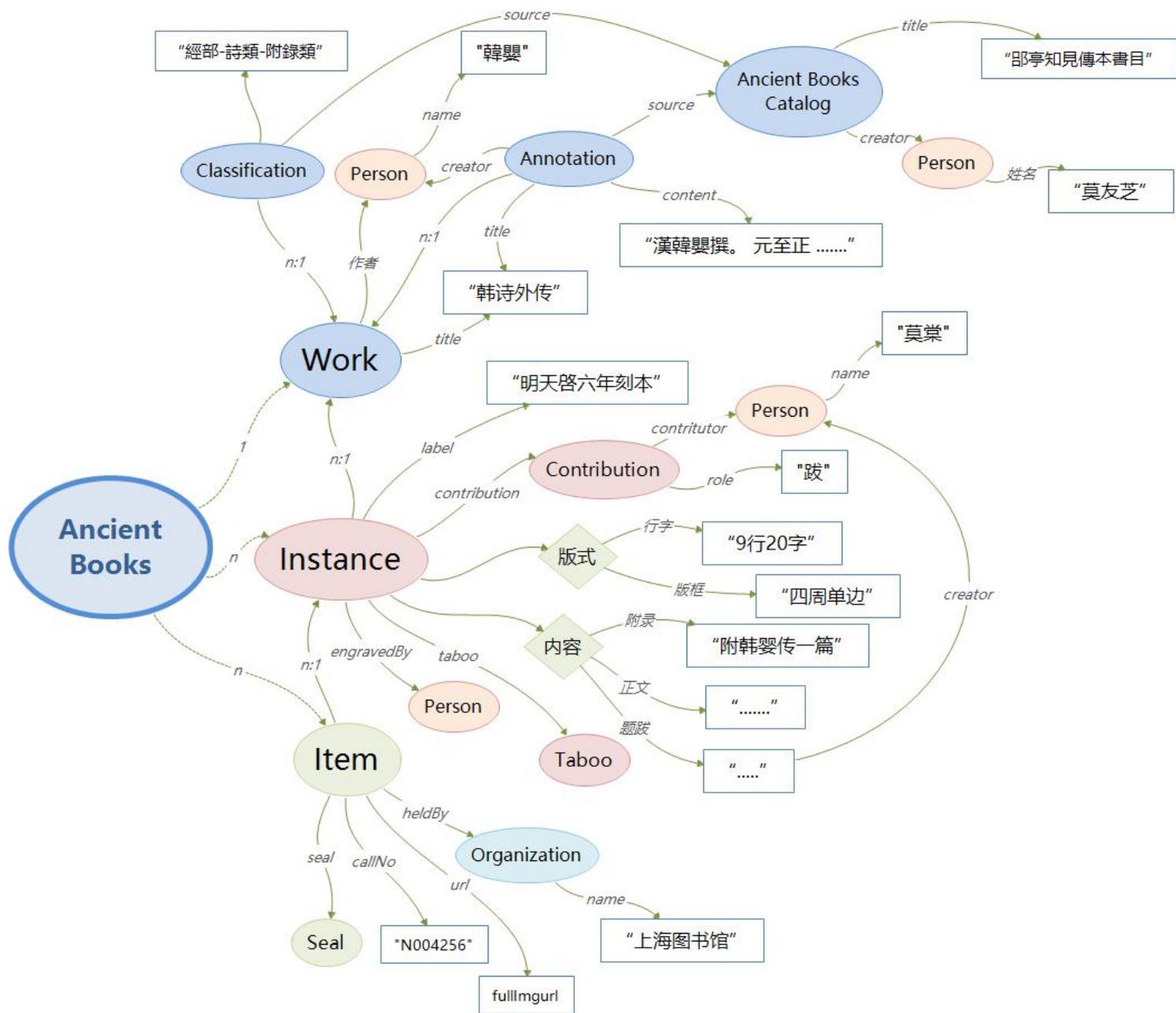
[http://pro.europeana.eu/files/Europeana\\_Professional/Share\\_your\\_data/Technical\\_requirements/EDM\\_Documentation//EDM\\_Primer\\_130714.pdf](http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Primer_130714.pdf)

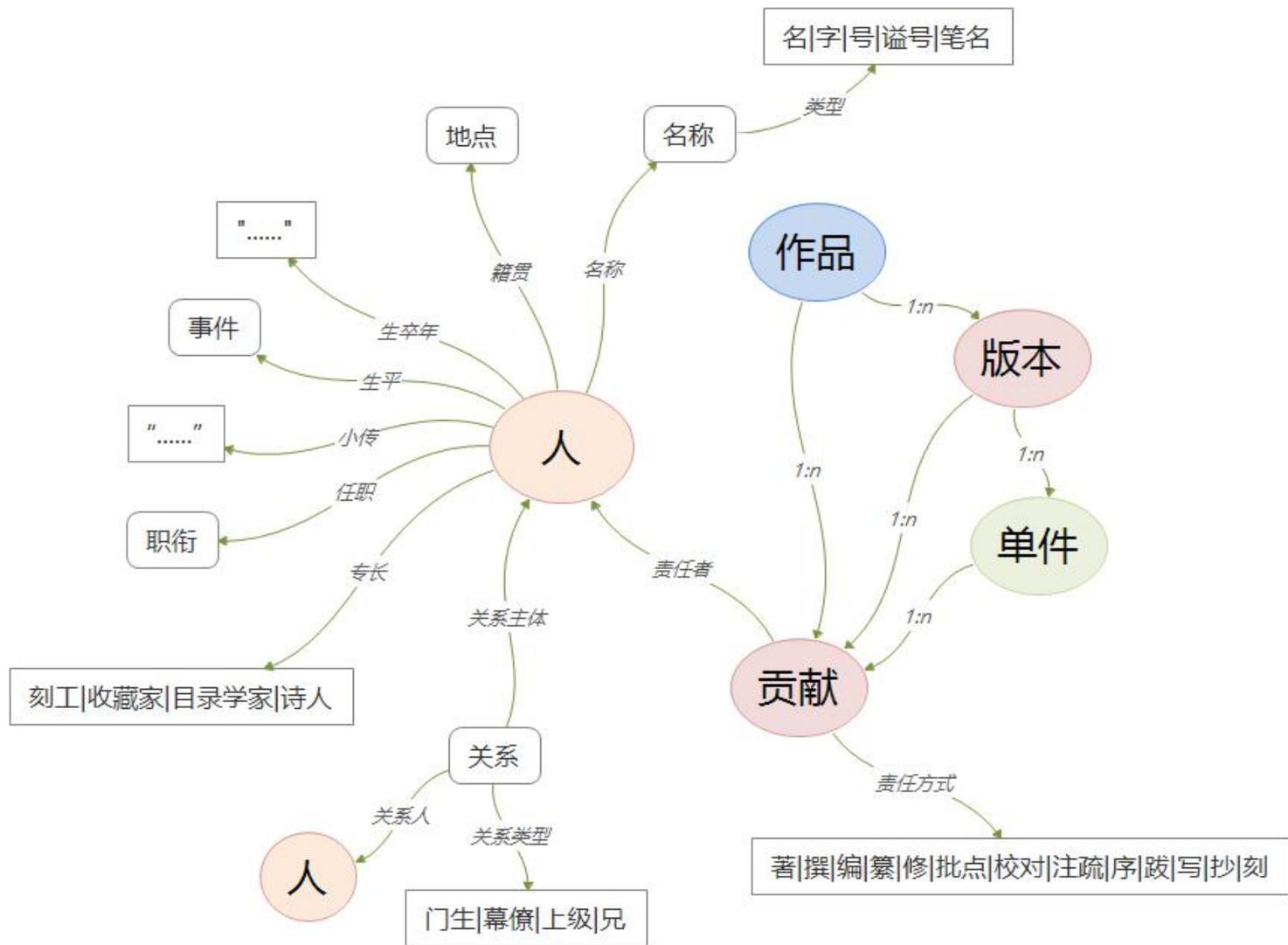
# 古籍数据建模：本体设计

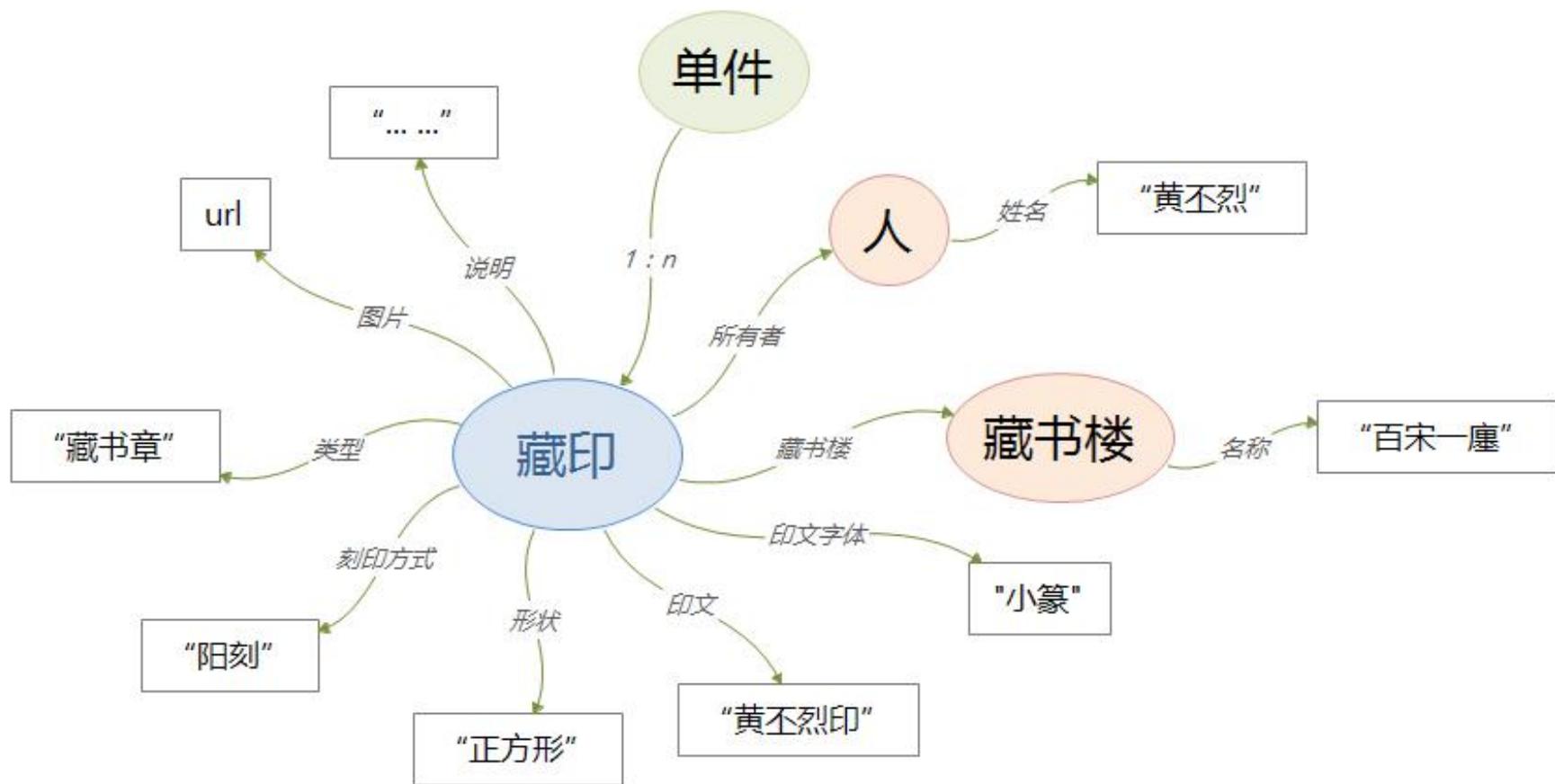
## Data Modeling: Ontology Design

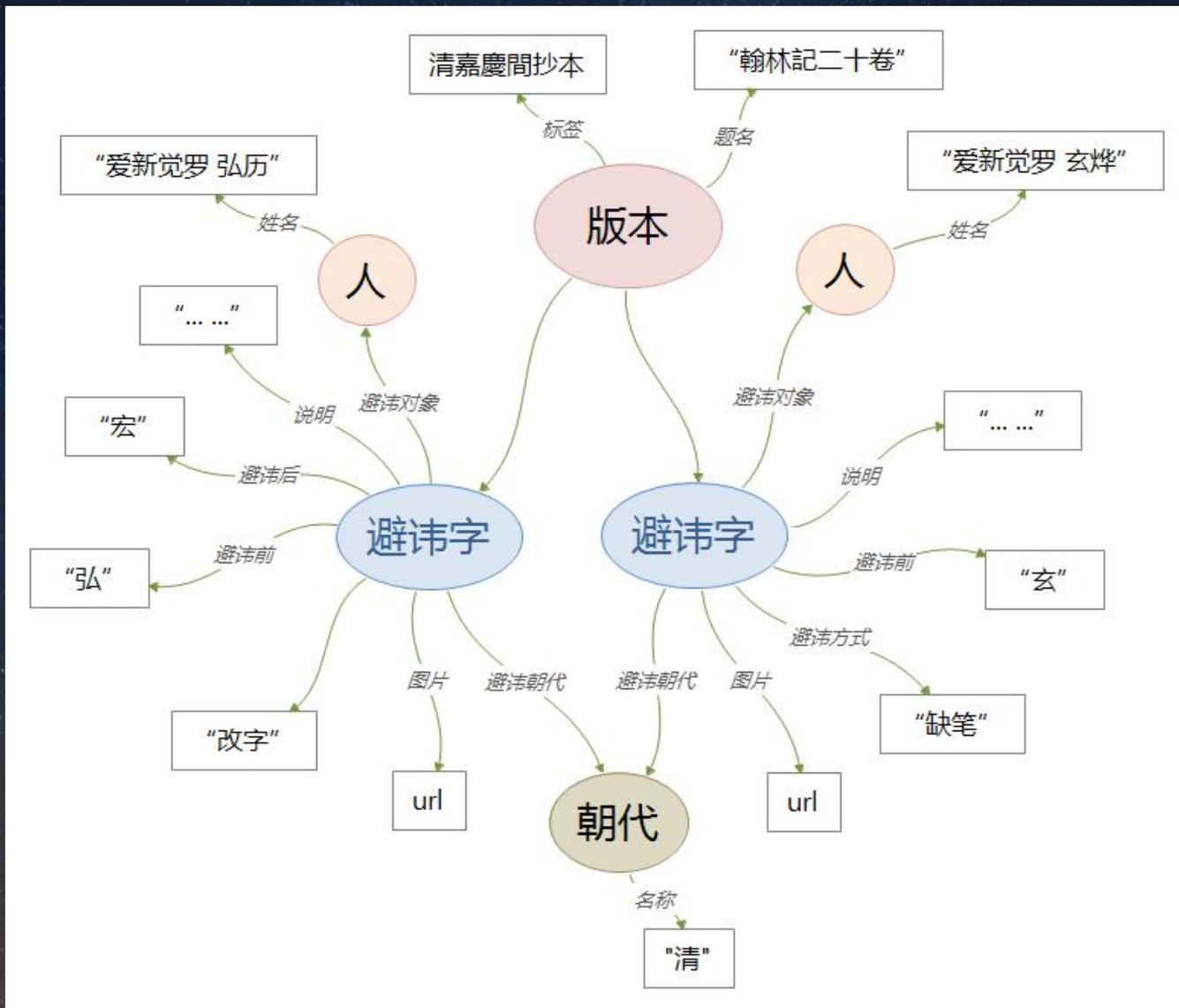
<http://gi.library.sh.cn/ontology>











# 古籍分类

古籍

分类1

宋史艺文志

分类2

贩书偶记

分类3

中国古籍善本  
书目

分类4

中国古籍总目

# 本体词表

gj.library.sh.cn/ontology/view/list

本页面视图展示了一个有序的分类表和属性列表。浏览器的后退或前进操作将调整页面上的导航位置。

## Classes

选择下面的一个类名并跳转到一个类面板和描述 其用法如下:

<a href="#">bf:Agent</a>	<a href="#">bf:Place</a>	<a href="#">pmb:Collection</a>	<a href="#">shl:Person</a>
<a href="#">bf:Category</a>	<a href="#">bf:Resource</a>	<a href="#">pmb:Instance</a>	<a href="#">shl:Place</a>
<a href="#">bf:Classification</a>	<a href="#">bf:Temporal</a>	<a href="#">pmb:Item</a>	<a href="#">shl:Resource</a>
<a href="#">bf:Collection</a>	<a href="#">bf:Topic</a>	<a href="#">pmb:Taboo</a>	<a href="#">shl:Temporal</a>
<a href="#">bf:Contribution</a>	<a href="#">bf:Work</a>	<a href="#">pmb:Work</a>	<a href="#">time:Instant</a>
<a href="#">bf:Identifier</a>	<a href="#">foaf:Person</a>	<a href="#">schema:Organization</a>	<a href="#">time:Interval</a>
<a href="#">bf:Instance</a>	<a href="#">gn:Feature</a>	<a href="#">schema:Place</a>	<a href="#">time:ProperInterval</a>
<a href="#">bf:Item</a>	<a href="#">org:Organization</a>	<a href="#">shl:Agent</a>	
<a href="#">bf:Organization</a>	<a href="#">pmb:Annotation</a>	<a href="#">shl:FamilyName</a>	
<a href="#">bf:Person</a>	<a href="#">pmb:Classification</a>	<a href="#">shl:Organization</a>	

## Properties

选择下面的一个属性名并跳转到一个属性面板和描述 其用法如下:

<a href="#">bf:agent</a>	<a href="#">foaf:name</a>	<a href="#">pmb:prefaceAndPostscript</a>	<a href="#">shl:dynasty</a>
<a href="#">bf:category</a>	<a href="#">gn:parentADM1</a>	<a href="#">pmb:seal</a>	<a href="#">shl:endDay</a>
<a href="#">bf:categoryType</a>	<a href="#">gn:parentADM2</a>	<a href="#">pmb:setName</a>	<a href="#">shl:endMonth</a>
<a href="#">bf:categoryValue</a>	<a href="#">gn:parentADM3</a>	<a href="#">pmb:spine</a>	<a href="#">shl:endYear</a>
<a href="#">bf:classification</a>	<a href="#">gn:parentADM4</a>	<a href="#">pmb:structureOfBook</a>	<a href="#">shl:genealogyName</a>
<a href="#">bf:contribution</a>	<a href="#">gn:parentCountry</a>	<a href="#">pmb:tableOfContents</a>	<a href="#">shl:generationCharacter</a>
<a href="#">bf:contributor</a>	<a href="#">oa:annotatedAt</a>	<a href="#">pmb:taboo</a>	<a href="#">shl:hasDataset</a>
<a href="#">bf:creator</a>	<a href="#">oa:annotatedBy</a>	<a href="#">pmb:tabooCharactor</a>	<a href="#">shl:hseb</a>
<a href="#">bf:dimensions</a>	<a href="#">oa:hasBody</a>	<a href="#">pmb:tabooPattern</a>	<a href="#">shl:level</a>
<a href="#">bf:edition</a>	<a href="#">org:headOf</a>	<a href="#">pmb:tabooTemporal</a>	<a href="#">shl:matieral</a>
<a href="#">bf:extent</a>	<a href="#">org:memberOf</a>	<a href="#">pmb:titlePage</a>	<a href="#">shl:monarch</a>

# 多源中文古籍数据融合

Multiple Source Data Mashup for Chinese Ancient Books

史志目录

官修目录

私家目录

藏书楼目录

现代联合目录

机构馆藏目录

表5 14种不同来源的古籍目录数据融合情况

类别	数据融合前		数据融合后		
	题名	书目种数	版本数量	注释数量	作品数量
史志目录	汉书·艺文志	614		614	290 005
	隋书·经籍志	3 168		3 168	
	旧唐书·经籍志	2 974		2 974	
	新唐书·艺文志	5 247		5 247	
	宋史·艺文志	9 188		9 188	
	明史·艺文志	3 673		3 673	
	清史稿·艺文志	7 176		7 176	
官修目录	崇文总目	3 389		3 389	
	四库全书总目	10 249		10 249	
私家目录	郎亭知见传本书目	3 681		3 681	
联合目录	中国古籍善本书目	约 6 万	96 728		
	中国古籍总目	约 20 万	407 581		
机构藏书目录	柏克莱加州大学 东亚图书馆中文 古籍善本书志	802	802	802	
	上海图书馆古籍数据库 (善本)	11 125	11 125		

# 联合目录的功能

## The Different Function Modules of the Platform

联合目录  
Union  
Catalogue  
for 1400+  
Organizations

合作机构单  
馆馆藏网站  
Website for  
one Org's  
Catalogue

古籍目录  
Evidence-  
based  
research  
(raw data)

循证研究  
Evidence-  
based  
research  
(Linked Data  
&tools)

中文古籍聯合目錄及循證平台

Chinese Ancient Books Union Catalogue and Evidence-based Platform

# 联合目录

## The Union Catalogue

標題	在此輸入標題	版本類型	在此輸入版本類型	版本時間	在此輸入版本時間	責任者	在此輸入責任者
批校	在此輸入批校	序跋	在此輸入序跋	出版地	在此輸入出版地	制作方法	在此輸入制作方法
冊數	在此輸入冊數	附註	在此輸入附註	尺寸	在此輸入尺寸	象鼻	在此輸入象鼻
魚尾							

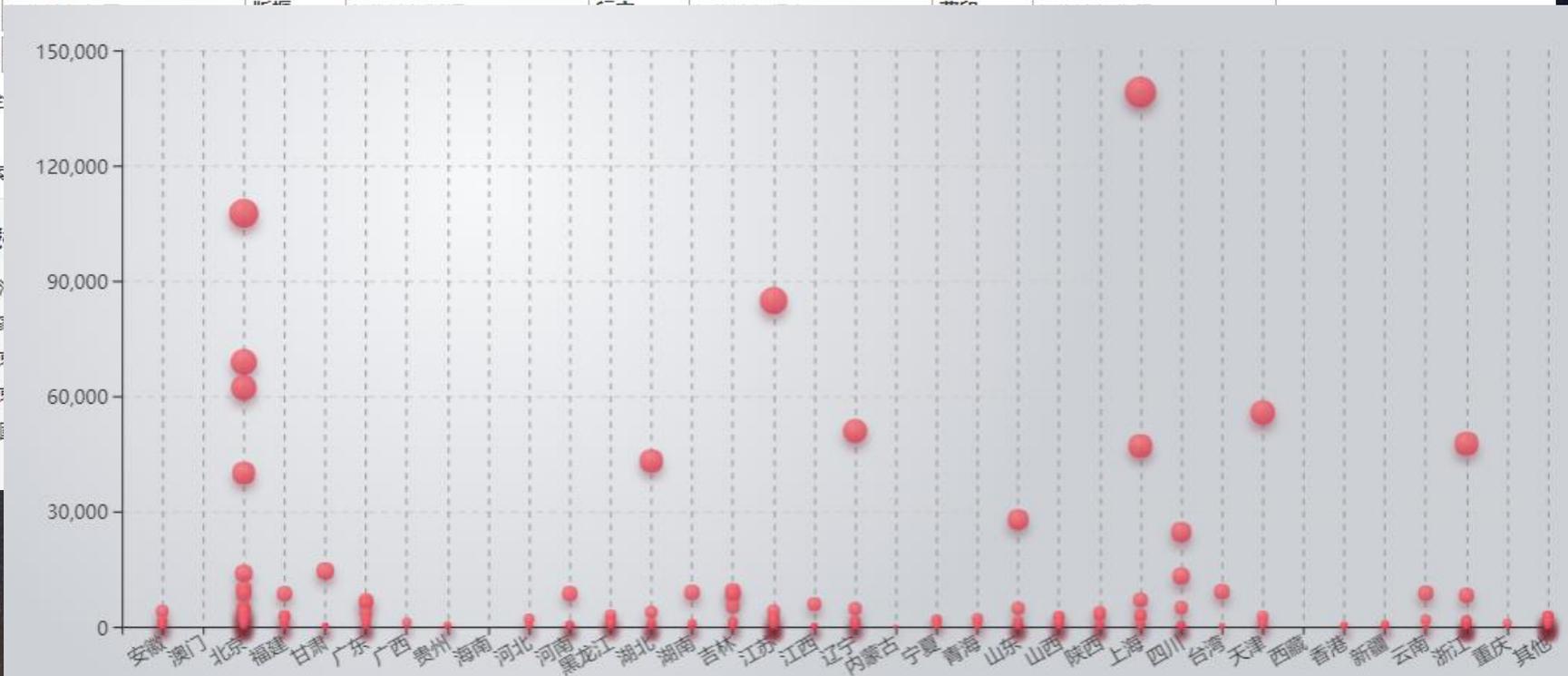
避諱字

上圖全

根據檢索

館藏

上海  
國  
南京  
北京  
中國



# 合作机构管理

## Website Configuration & Data Manage for Organizations

### 合作机构管理

#### 机构馆藏管理

- 本馆馆藏目录认定
- 本馆馆藏目录导出
- 馆藏目录查询修改
- 馆藏目录批量上传

#### 机构主页管理

- 默认检索方式
- 高级检索
  - 简单检索

#### 高级检索字段

- 分类
- 版本类型
- 尺寸
- 避讳方式

#### 显示分面

- 标题
- 版本时间
- 藏印
- 避讳朝代

- 责任者
- 出版地
- 象鼻
- 索书号

#### 详细页显示字段

- 批校
- 制作方法
- 鱼尾
- DOI

#### 排序字段

- 序跋
- 板框
- 册数
- 全文

- 附注
- 行字
- 避讳字

### 中文古籍联合目录及循证平台

Chinese Ancient Books Union Catalogue and Evidence-based Platform

- 联合目录
- 循证研究
- 古籍目录
- 个人中心
- 册数统计
- 数据维护
- 合作机构管理
- 系统管理

#### 美国加州大学柏克莱分校东亚馆

站点标题

Logo  未选择任何文件

预览 

站点图标  未选择任何文件

预览 

模板  

馆藏集合名称

# 合作机构——定制的单馆馆藏网站

Customized Website for One Organization's Collection Database



澳門大學  
UNIVERSIDADE DE MACAU  
UNIVERSITY OF MACAU

🏠 首页

☰ 联合目录

U 循证研究

📖 古籍目录

👤 登录

➔ 注册

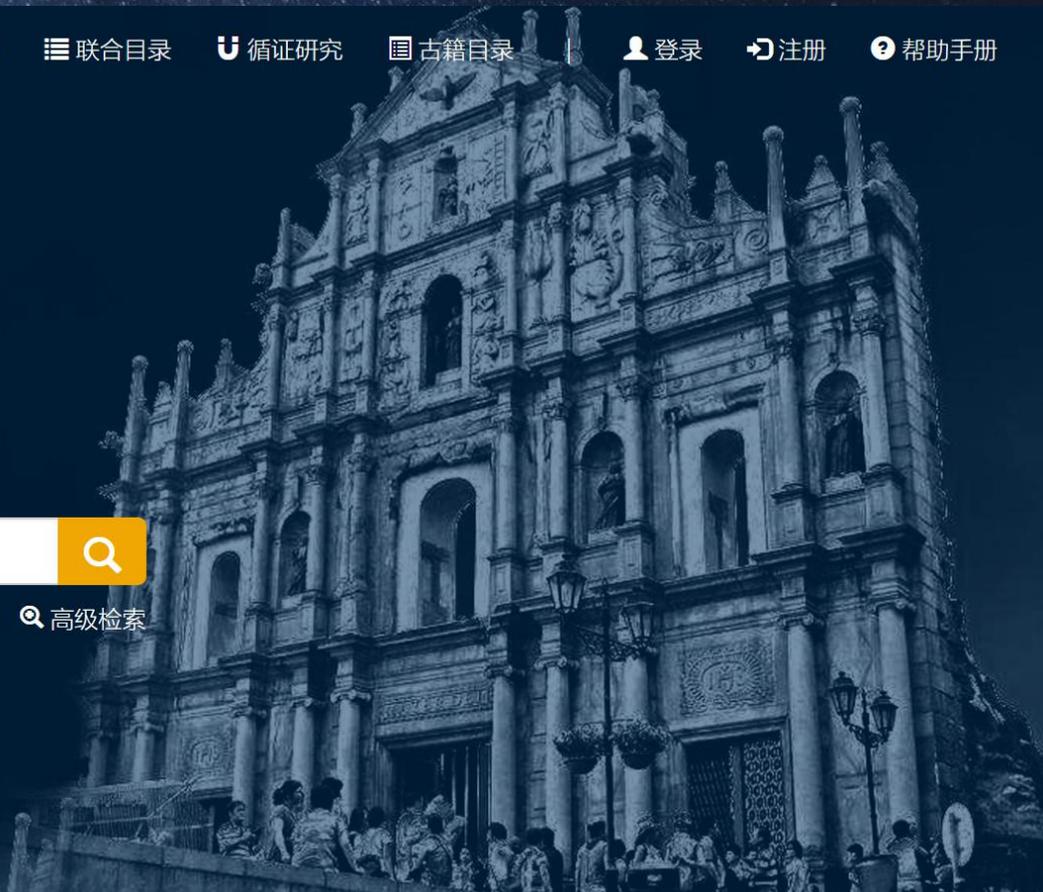
📖 帮助手册

## 澳门大学图书馆 中文古籍目录

题名/責任者/批校題跋者/版本類型



🔍 高级检索



# 合作机构——定制的单馆馆藏网站

Customized Website for One Organization's Collection Database



🏠 首页

☰ 联合目录

U 循证研究

📖 古籍目录

👤 登录

🔄 注册

📖 帮助手册

## 上海图书馆古籍目录

SHANGHAI LIBRARY ANCIENT BOOKS CATALOGUE

题名/责任人/批校题跋者/版本类型



🔍 高级检索

# 合作机构——定制的单馆馆藏网站

Customized Website for One Organization's Collection Database

Berkeley  
UNIVERSITY OF CALIFORNIA

🏠 首页

☰ 联合目录

U 循证研究

📖 古籍目录

👤 登录

➔ 注册

📖 帮助手册

## 伯克莱加州大学东亚图书馆 汉文古籍目录

题名/责任者/批校题跋者/版本类型



🔍 高级检索

# 合作机构——定制的单馆馆藏网站

Customized Website for One Organization's Collection Database

**HARVARD UNIVERSITY**

🏠 首页    📖 联合目录    🔍 循证研究    📚 古籍目录    👤 登录    📄 注册    📖 帮助手册

## 哈佛燕京图书馆藏 中文善本古籍特藏系统

题名/责任人/批校题跋者/版本类型

🔍 高级检索

# 古籍目錄

## Ancient Catalogue

漢書

隋書

部

善中

### 《四庫全書總目提要》

題名：

責任者：

內容：

根據檢索條件為您檢索到10249條記錄

#### 分類

- 經(1776)
- 史(2133)
- 子(2941)
- 集(3399)

#### 子夏易傳

舊本題「卜子夏撰」。案說《易》之家，最古者莫若是書。其偽中生偽，至一至再而未已者，亦莫若是書。《唐會要》載開元七年詔：“子夏《易傳》，近無習者，令儒官詳定。”劉向《七略》曰：“《漢志》《易》有十三家而無子夏作傳者，至梁阮氏《七錄》，始有《子夏易》六卷，或云韓嬰作，或云丁寬作。然據《漢書》，《周易》十二篇，《丁易》八篇，求其符合，事殊確刺，必欲行用，深以為疑。”司馬貞議亦曰“案劉向《七略》有《子夏易傳》，但此書不行已久，今所存多失真本。荀勗《中經簿》云：《子夏傳》四卷，或云丁寬。是先違疑非子夏矣。又《隋書·經籍志》云：《子夏傳》殘闕，梁六卷，今二卷。知其書錯繆多矣。又王儉《七志》引劉向《七略》云：《易傳》子夏，韓氏嬰也。今題不稱韓氏而載薛虞記，其質粗略，旨趣非遠，無益后學”云云。是唐以前所謂《子夏傳》，已為偽本。晁說之《傳易堂記》又稱：“今號為《子夏傳》者，乃唐張弧之《易》”（案弧唐末為大理寺評事，有《素履子》，別著錄），是唐時又一偽本并行。故宋《國史志》以《假托子夏易傳》與《真子夏易傳》兩列其目，而《崇文總目》亦稱此書篇第，略依王氏，決非卜子夏之文也。朱彝尊《經義考》，證以陸德明《經典釋文》、李鼎祚《周易集解》、王應麟《困學紀聞》所引，皆今本所無。德明、鼎祚猶曰在張弧以前，應麟乃南宋末人，何以當日所見與今本又異？然則今本又出偽托，不但非子夏書，亦并非張弧書矣。流傳既久，姑存以備一家云爾。（謹案：唐徐堅《初學記》以太宗御制升列歷代之前，蓋臣子尊君之大義。焦竑《國史經籍志》、朱彝尊《經義考》并踵前規。臣等編摩《四庫》，初亦恭錄御定《易經通注》、御纂《周易折中》、御纂《周易述義》弁冕諸經。仰蒙指示，命移冠國朝著述之首，俾尊卑有序而時代不淆。聖度謙沖，酌中立憲，實為千古之大公。謹恪遵彝訓，仍托始於《子夏易傳》，并發凡於此，著《四庫》之通例焉。又案：托名之書有知其贗作之人者，有不知其贗作之人者，不能一一歸其時代，故《漢書·藝文志》仍從其所托之時代為次。今亦悉從其例。）



谢谢THANKS!

[cjxia@libnet.sh.cn](mailto:cjxia@libnet.sh.cn)