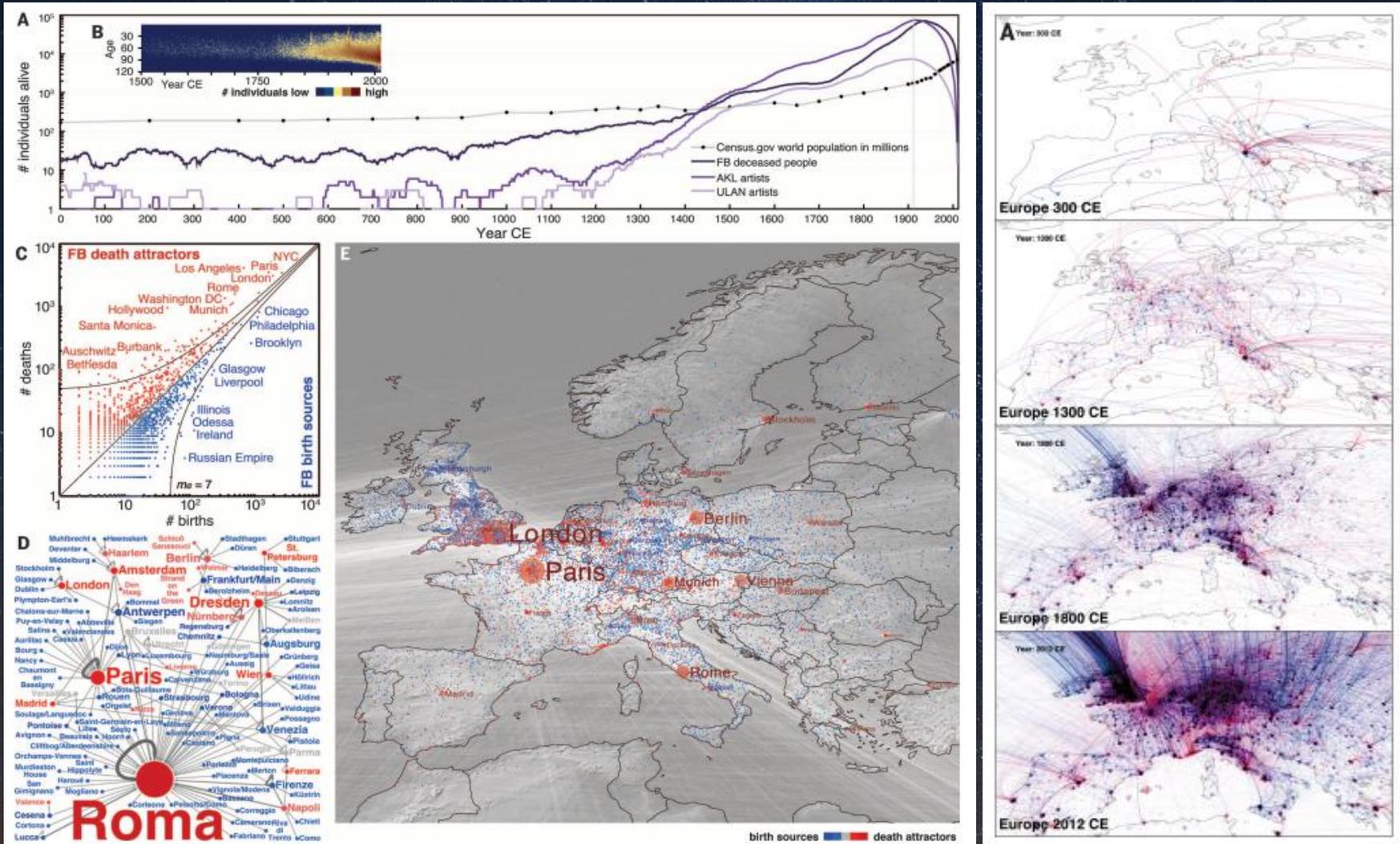


# 数字人文中的人名规范控制

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

# “人” 是人文研究的重要因素



Schich M, Song C, Ahn Y Y, et al. A network framework of cultural history[J]. Science, 2014, 345(6196):558-562.

# 作为人文研究的基础设施

## 中國歷代人物傳記資料庫

中央研究院歷史語言研究所 · 哈佛大學 · 北京大學中國古代史研究中心合作開發  
Developed through collaboration among Academia Sinica, Harvard University, and Peking University

人名查詢 地名查詢 官名查詢 關鍵詞查詢 進階查詢 親屬/人際關係查詢 檢索結果輸出 社會關係分析 操作說明 名詞說明

人名查詢 2 4 查詢

查詢結果 5筆

傳主資料

姓名/別名/外文名: 王維 1

年代 (指數年): 0618 至 0907

朝代: Tang 唐 (618-907) 3

地名: [尚未定義]

基本資料查詢結果(傳主) 5筆

檢索條件: 共5筆, 第1/1頁, 自第1至第5筆, 每頁顯示 20 筆

序號	Name	Dynasty	Choronym	Alternate Names	外文名
1	He Zhizhang 賀知章	唐Tang	Kuayuan 會稽	He jian賀監(Unknown未詳); Jizhen季真(Courtesy name字); Simingkuangke 四明狂客(Unknown未詳); Weimo維摩(Courtesy name字)	7
2	Cheng Wei 程偉	唐Tang	Xinan 信安		8
3	Liu Hun 柳渾	唐Tang	Hedong 河東	Liu Ba 八(Birth-order name行第); Liu Hun 柳渾(Alternate Personal Name, Previously Used Name別名、曾用名); Liu Zai 棗(Alternate Personal Name, Previously Used Name別名、曾用名); Weishen 諱深(Courtesy name字); Yikuang 彙(Courtesy name字); Yuanyu 元虞(Courtesy name字); Zhen 真(Posthumous Name諡號)	7
4	Wei Wei 魏維	唐Tang	Jingzhao 京兆		7
5	Wang Wei 王維	唐Tang	Hedong 河東	Moje 摩詰(Courtesy name字); Wang Shisan 王十三(Birth-order name行第); Wang youcheng 王右丞(Unknown未詳)	

## 百家諸子 Chinese Text Project

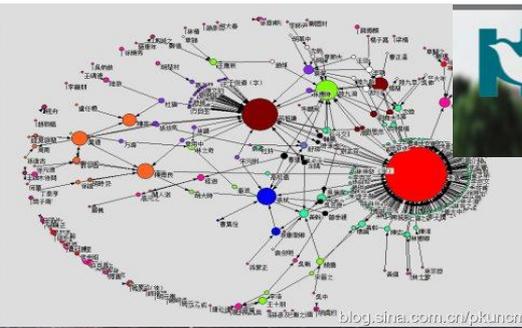
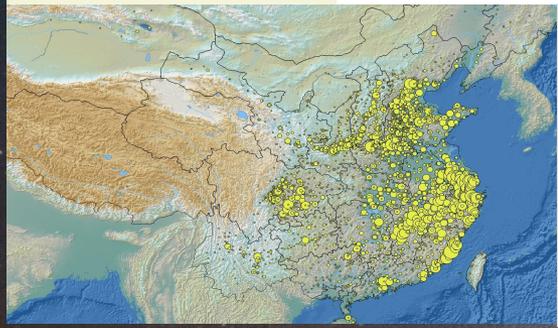
MARKUS

MARKUS

你可以使用码库思, 为你上传的中文档案, 自动标记人名、地名、年号及官名(日后也可应用在其他语言), 你也可以自订要标记的关键词。标记过后, 你可以边阅读一篇文章, 边参考相关文献, 或比较两段标记相同关键词的段落。你也可以提取你的标记信息, 以我们的视觉化平台或其他工具, 作进一步的分析。

第一步: [上传 UTF-8 纯文字或 MARKUS 档案](#) 或 [把文字贴在这里](#)

MARKUS was developed as part of the project "Communication and Empire: Chinese Empires in Comparative Perspective," funded by the European Research Council. Further development has been funded by DfID, JISC and AHRC. Code licensed AGPL v3, docs and videos CC BY SA 4.0.



## K&SS 学术地图发布平台

Academic Map Publishing Platform

<https://projects.iq.harvard.edu/cbdb/home>

什么是规范控制？



[美] 埃雷兹·艾登 / [法] 让-巴蒂斯特·米歇尔. 可视化未来：数据透视下的人文大趋势

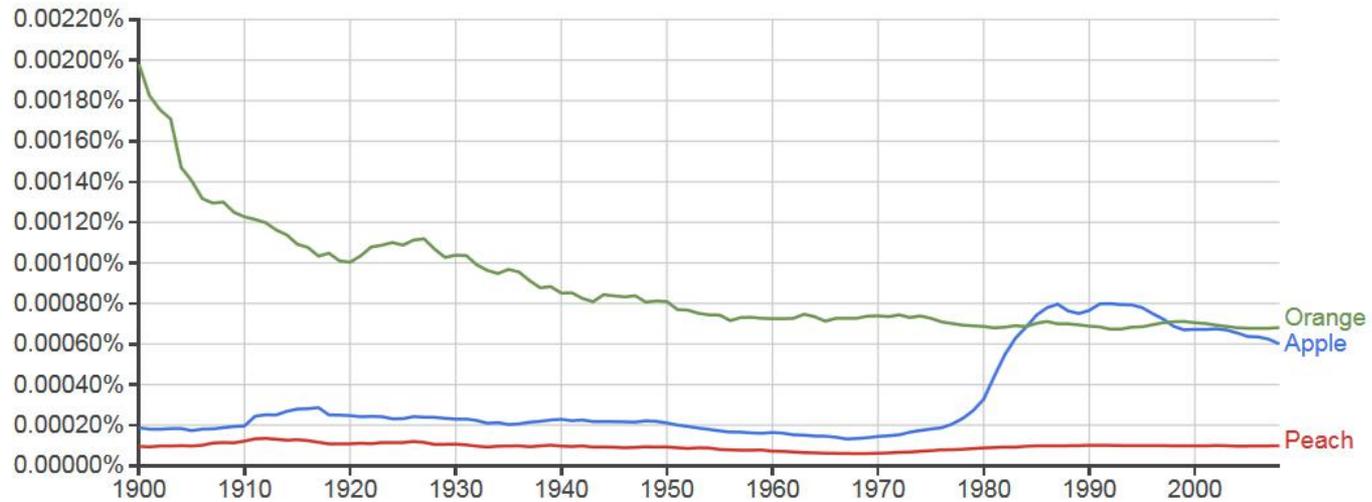
Uncharted: Big Data as a Lens on Human Culture

## Google Books Ngram Viewer

Graph these comma-separated phrases:   case-insensitive

between  and  from the corpus  with smoothing of

[Search lots of books](#)



# Google Books Ngram Viewer

Graph these comma-separated phrases:

case-insensitive

between  and  from the corpus

with smoothing of

[Search lots of books](#)



# 人名规范控制

同名消歧

异名合并

# 图书馆的人名规范控制

“集中同一责任者的不同著作，汇集同一著作的不同版本或译本，提高文献检索率”——丘东江.图书馆学情报学大辞典

国际图联 FRISAD (2008)  
国际图联 LRM (2017)  
美国国会图书馆 (2016)  
RDA MARC21

检索点

标识符

字符串 (String)

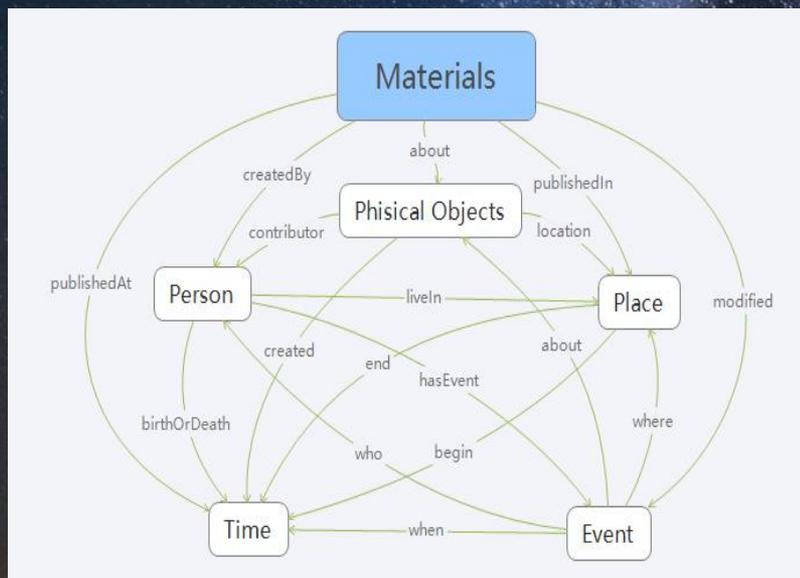
实体 (Entity)



# 数字人文中的人名规范控制

## 知识组织

## 数据服务



### 人名

蒲牢 雁冰 方璧 有茅盾 沈德鸿 沈鸿 沈雁冰 燕昌 德鸿 瘦沈郎  
玄珠 形天 亦作保中 沈仲方 明甫 冰师 仲方 仲芳 沈明甫  
方保宗 方保中 毛姑 矛盾 小凡 凡 山石 子荪 子渔 子敬 乡愚  
韦 韦兴 韦君 元枚 云 止水 止敬 水 忆秋生 毛腾 公羽 风  
孔常 文 文直 方 方非 方璧 未名 未明 石萌 石崩 丙 丙申  
世珍 丙申生 丙生 平 东方未明 冬芬 叶明 四珍 玄 玄瑛 冯夷

### 关系

- |         |        |        |
|---------|--------|--------|
| 孔德沚(妻)  | 曹辛之(友) | 黎澍(友)  |
| 赵清阁(友)  | 姚雪垠(友) | 夏衍(友)  |
| 刘白羽(友)  | 萨空了(友) | 乐黛云(友) |
| 端木蕻良(友) | 熊佛西(友) | 楼适夷(友) |

# 规范档Vs. 人物资料库

	目的	人物选取	内容组织	建置方法	建置技术
规范档	同一作者的作品聚类	文献的责 任者	规范名、异 名、生卒年、 著述	图书馆员在编 目工作中遵循 行业标准逐步 增加	关联开放 数据
人物资料 库	人物资料 的全面收 集	全面地收 集某一个 领域（如 诗人）或 某一个时 代（如明 清）中的 所有人物	姓名、生卒 年、异名、 籍贯、传略、 出身、专长、 著述、社会 关系、经历、 官职、传记 资料	相关领域的研 究人员收集整 理，具有较强 的领域性和专 业性	关系数据 库

以上图人名规范库为例

# 上图数字人文平台

## Shanghai Library's DH Platform

### 中文古籍聯合目錄及循證平台

Chinese Ancient Books Union Catalogue and Evidence-based Platform

<http://gj.library.sh.cn>



上海历史文化时空再造  
——从武康路出发

<http://wkl.library.sh.cn>

上海图书馆

名人手稿档案库

<http://sg.library.sh.cn>

上海市文献联合编目中心  
关联书目数据发布平台

<http://bib.library.sh.cn>

家譜 知識服務平台  
Genealogy knowledge service platform

<http://jiapu.library.sh.cn>

盛宣怀档案知識庫

<http://sd.library.sh.cn>

规范控制

Sparql Endpoint+ API +HTTP URI

<http://names.library.sh.cn>

地

时

人名  
规范库  
843493

马路

建筑

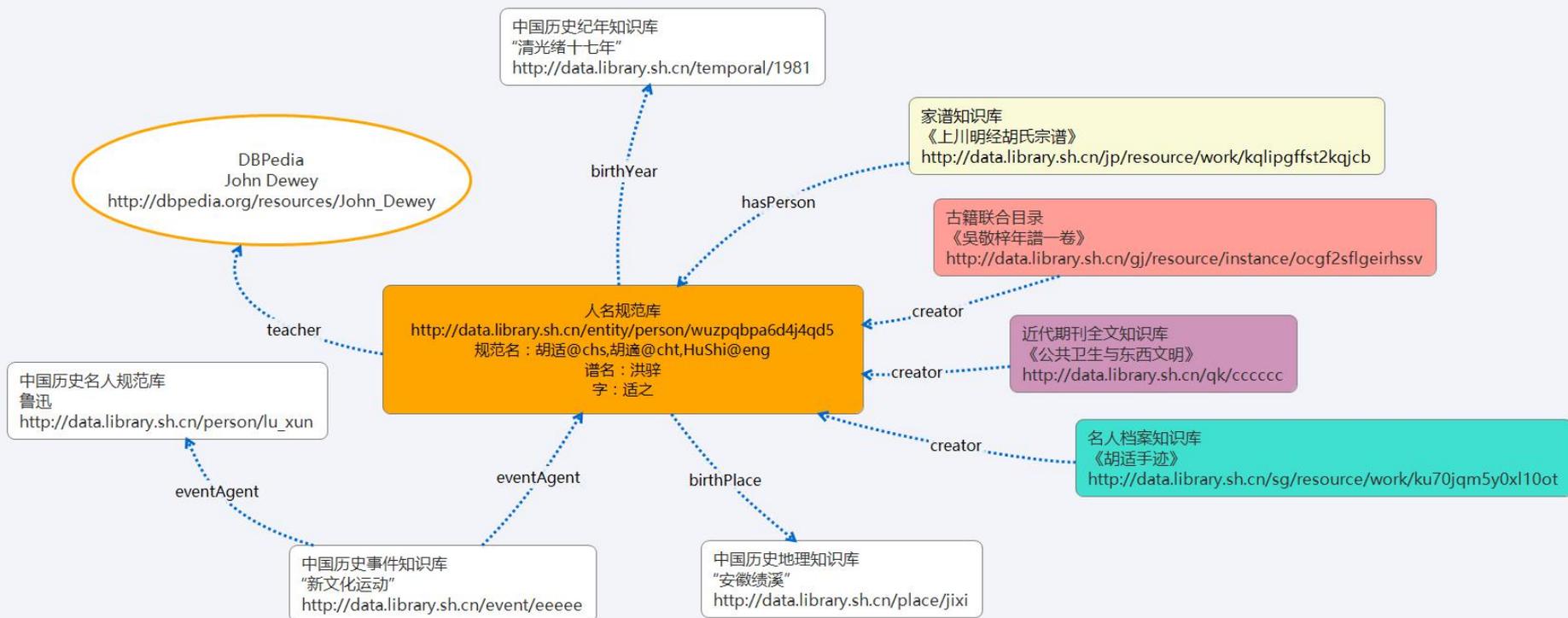
事件

<http://data.library.sh.cn> as **Data Infrastructure**

# 基于万维网的人名规范控制

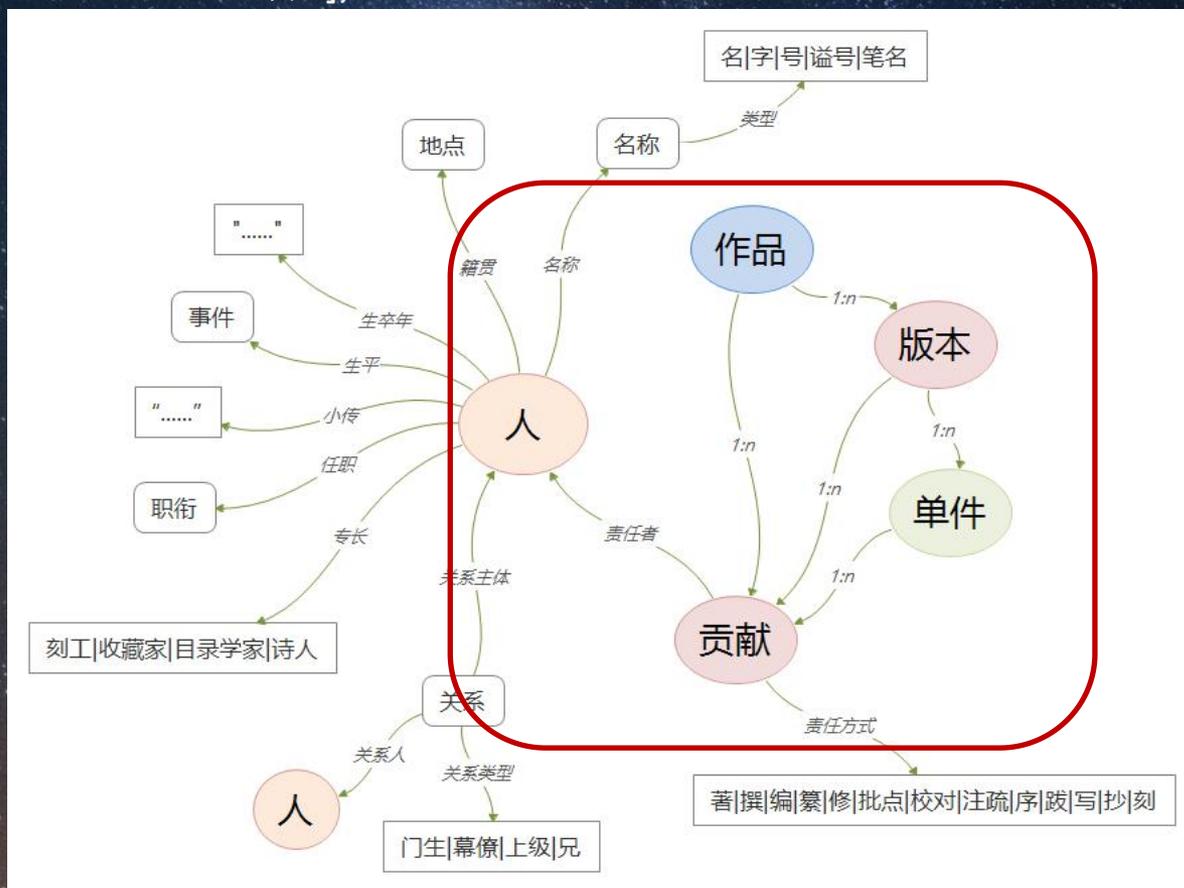
## Names Authority Control on the Web

- A person is an **entity** has a **HTTP URI** which is not only a unique **identifier** but also a universal **locator** on the Web.
- If different documents in different namespaces have the same creator, the value of *creator* is referenced to the same unique HTTP URI of the person.



# 人与文献的关联——以古籍为例

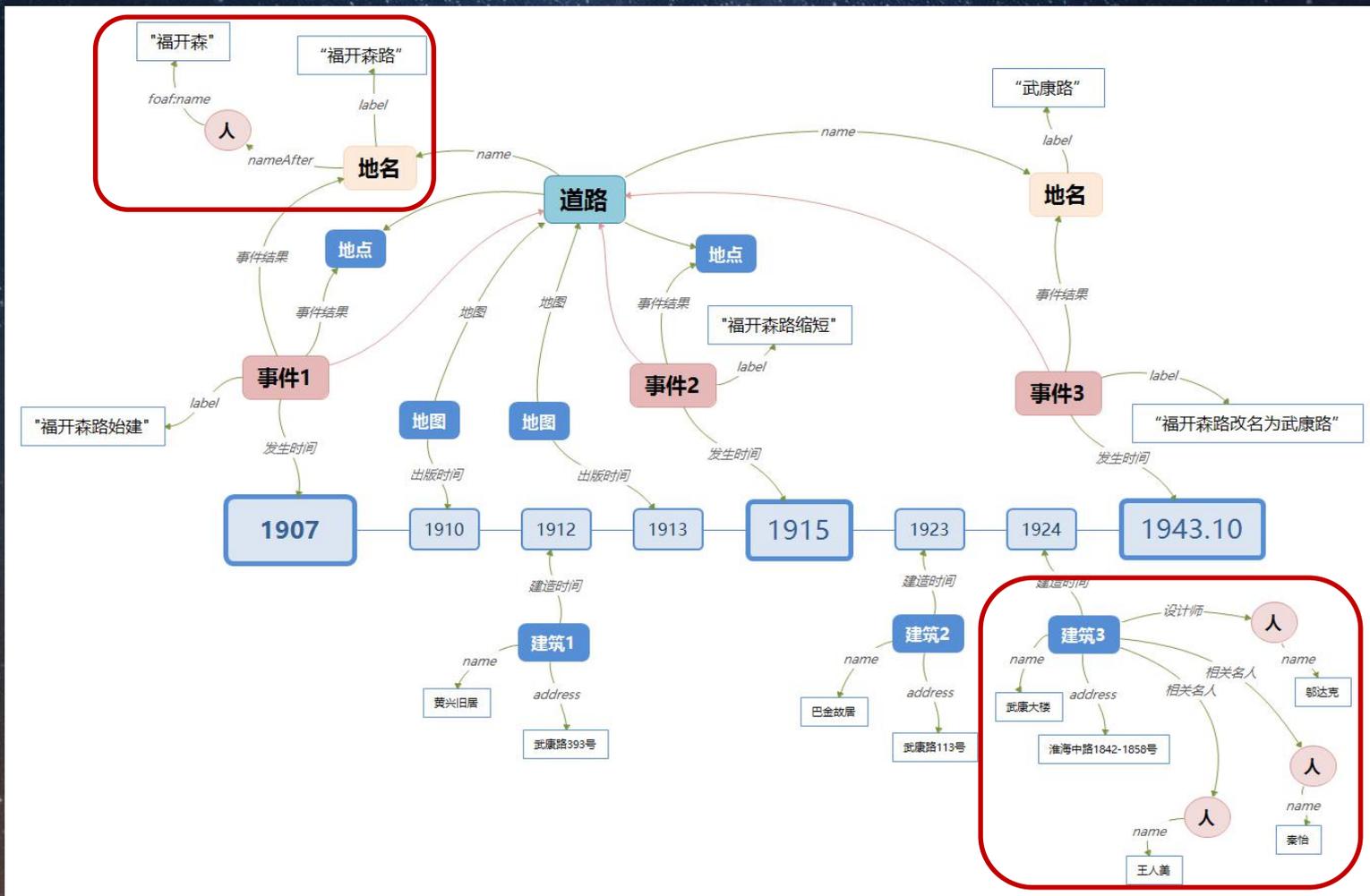
<<http://data.library.sh.cn/gj/resource/instance/m6ebrlz45xkazx14>> a pmb:Instance;  
bf:contribution [ a bf:Contribution;  
bf:creator <<http://data.library.sh.cn/entity/person/g6fw16h6ky7jbrp3>>,  
bf:role “著”];



# 人与建筑的关联——以武康路为例

<<http://data.library.sh.cn/entity/architecture/amknmwvk01qaykng>> shl:designer

<<http://data.library.sh.cn/entity/person/e8zbuk7w0ud3ztd6>>





# 开放的、可重用的、可共享的本体词表

## Open, reusable and shareable vocabulary

```

122 shl:Person a rdfs:Class ;
123     rdfs:label "人物" ;
124     rdfs:subClassOf foaf:Person ;
125     rdfs:subClassOf bf:Person ;
126     rdfs:subClassOf shl:Agent ;
127     rdfs:comment "人物。" .
128
129 shl:FamilyName a rdfs:Class ;
130     rdfs:label "姓氏" ;
131     rdfs:subClassOf bf:Topic ;
132     rdfs:subClassOf shl:Resource ;
133     rdfs:comment "姓氏。" .
134
135 shl:Organization a rdfs:Class ;
136     rdfs:label "机构" ;
137     rdfs:subClassOf bf:Organization ;
138     rdfs:subClassOf schema:Organization ;
139     rdfs:subClassOf org:Organization ;
140     rdfs:subClassOf shl:Agent ;
141     rdfs:comment "机构。" .
142
143 shl:Place a rdfs:Class ;
144     rdfs:label "地点" ;
145     rdfs:subClassOf prov:Location ;
146     rdfs:subClassOf shl:Resource ;
147     rdfs:subClassOf gn:Feature ;
148     rdfs:subClassOf bf:Place ;
149     rdfs:comment "地点" .
150
151 shl:Temporal a rdfs:Class ;
152     rdfs:comment "朝代。";
153     rdfs:label "朝代";
154     rdfs:subClassOf shl:Resource ;
155     rdfs:subClassOf time:Interval

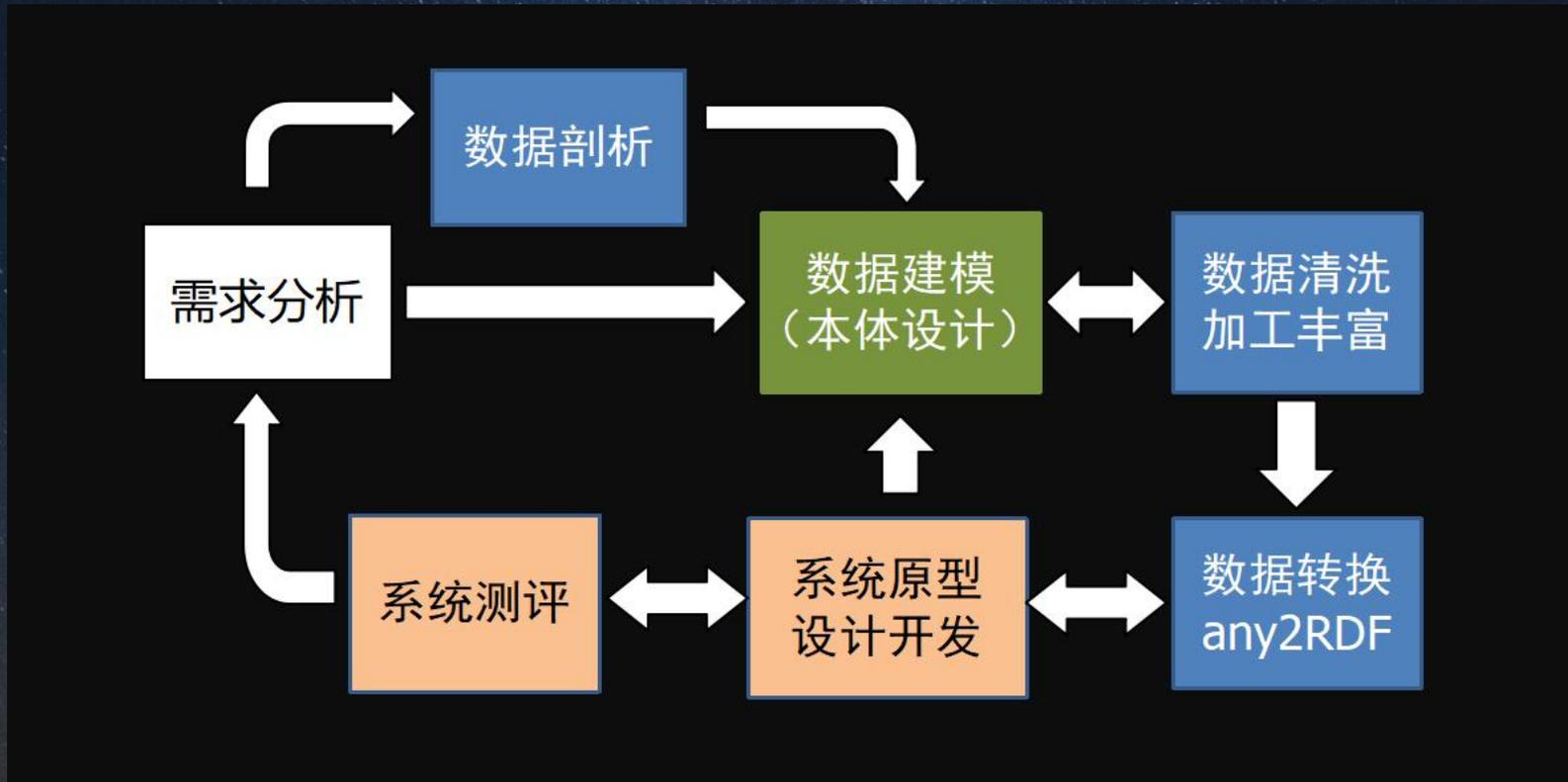
```

machine  
understandable

Property of	Label	Used With	Expected Value
	姓氏	shl:Person	shl:FamilyName
	性别	shl:Person	bf:Category
	父母	shl:Person	shl:Person
	朋友	shl:Person	shl:Person
	子女	shl:Person	shl:Person
	配偶	shl:Person	shl:Person
	生于	shl:Person	rdfs:Literal
	出生地	shl:Person	shl:Place
	小传	shl:Person	rdfs:Literal
	字	shl:Person	rdfs:Literal
	著作	shl:Person	rdfs:Literal
	著作	shl:Person	bf:Work
	卒于	shl:Person	rdfs:Literal
	死亡地	shl:Person	shl:Place
	民族	shl:Person	bf:Category
	谱名	shl:Person	rdfs:Literal
	字辈	shl:Person	rdfs:Literal

# 渐进式而非一步到位的建库流程

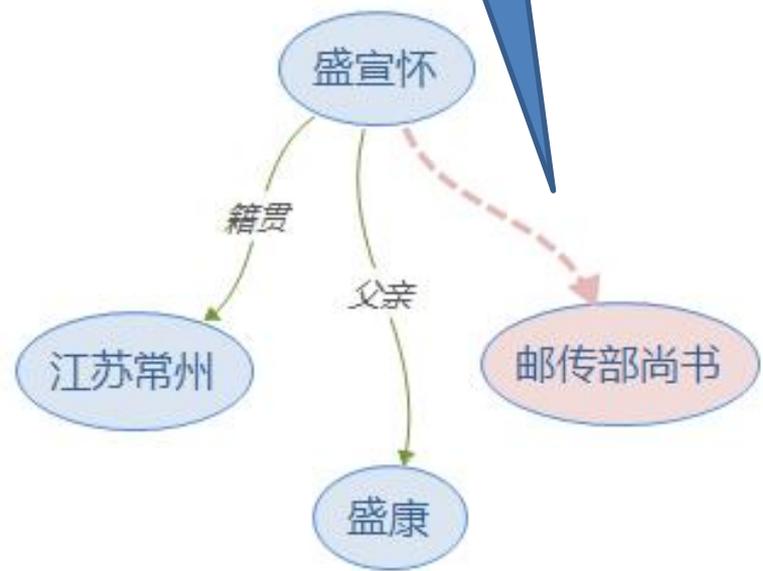
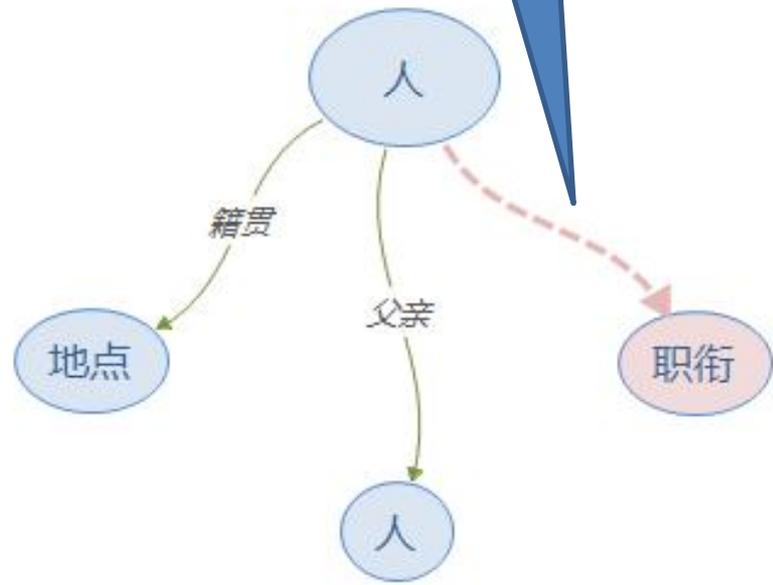
The workflow of building the knowledgebase



夏翠娟,张磊,贺晨芝.面向知识服务的图书馆数字人文项目建设:方法、流程与技术[J].图书馆论坛,2018,38(01):1-9.

增加本体属性

增加实例数据



# 数据情况

## Data Summary

1. 收集依据：图书馆馆藏文献中的作者，文献内容中涉及到的重要人物

The creators and contributors of the documents in libraries, or some important people mentioned in the content.

2. 数据来源：元数据、上图近现代人名规范档、国图规范档、网络百科、人名大辞典、CBDB

Data Source: Metadata, Shanghai Library's Authority control names (1800-1949), Encyclopedia, CBDB, Dictionary of names.

3. 数量：84万多 More than 840 thousands

## 朝代

当代(171624) 民国(52413) 南明(1)  
清(21446) 明(1779) 元(165)  
南宋(7) 北宋(1) 宋(816)  
五代(16) 唐(58) 隋(7)  
南北朝(15) 晋(12) 西晋(1)  
三国(2) 汉(53) 东周(1)  
周(19) 其他(595758)

## 籍貫

江苏(14035) 山东(12703) 浙江(12050)  
湖南(10744) 河南(8918) 湖北(8771)  
河北(7674) 安徽(6973) 四川(6096)  
广东(6059) 辽宁(5666) 江西(5219)  
山西(4939) 陕西(4518) 福建(4398)  
黑龙江(3024) 上海(2997) 吉林(2934)  
北京(2775) 甘肃(2659) 广西(2221)  
云南(2152) 重庆(2099) 天津(1573)  
内蒙古(1534) 贵州(1184) 台湾(913)  
宁夏(521) 海南(457) 新疆(375)  
青海(232) 香港(74) 西藏(49)  
澳门(6) 其他(697620)

# 如何作为数据基础设施

## How to provide data service as Data Infrastructure

1. 申请API Key  
data.library.sh.cn



2. 阅读接口说明文档  
data.library.sh.cn

接口  
调用

### 3.1. SPARQL Endpoint

特点:

Web 调用, 跨网域搜索, 要求精通 SPARQL 语言。

网址:

<http://data1.library.sh.cn:8890/sparql>

### 3.2. Restful API

特点: Web调用, 跨语言, 门槛低。

方法:

[http://data.library.sh.cn/sg/person/data?\[参数1\]&\[参数2\]&\[参数3\]?key=YourAPIKey](http://data.library.sh.cn/sg/person/data?[参数1]&[参数2]&[参数3]?key=YourAPIKey)

### 3.3. HTTP URI内容协商

特点: Web调用, 跨语言, 访问实体的 HTTP URI 即可获得其结构化数据。

方法: 需先通过调用前两种接口获得实体的 HTTP URI

返回数据  
格式:

**JSON-LD**

W3C推出的用于关联数据消费的JSON格式

# 如何作为数据基础设施: sparql endpoint

How to provid data service as Data Infrastructure

← → ↻ data.library.sh.cn:8890/sparql

Virtuoso SPARQL Query Editor [About](#) | [Namespace Pref](#)

Default Data Set Name (Graph IRI)

Query Text

```
select distinct ?Concept where {[[] a ?Concept} LIMIT 100
```

Sponging: Use only local data (including data retrieved before), but do not retrieve more ▼

Results Format: HTML ▼

Execution timeout: 0 milliseconds (values less than 1000 are ignored)

Options:  Strict checking of void variables

(The result can only be sent back to browser, not saved on the server, see [details](#))

Run Query Reset

Sparql  
Endpoints  
make The  
Web as a  
gloable  
database

# 如何作为数据基础设施: HTTP URI

## How to provide data service as Data Infrastructure

The screenshot shows the Shanghai Library website with a search bar and navigation tabs for '人物' (People) and '地点' (Locations). A profile for 茅盾 (Maodun) is displayed, including a bio and a list of friends. A red box highlights the 'RDF | NT | XML | JSON' links. A callout box points to the 'foaf:name' property in the JSON-LD data, with the text 'Machine understandable metadata Schema(Ontology)'. The JSON-LD data is as follows:

```
JSON-LD
{
  "http://purl.org/vocab/relationship/spouseOf" : [ {
    "type" : "uri",
    "value" : "http://data.library.sh.cn/entity/person/8gqhwuqdhstgrwgv"
  } ],
  "http://purl.org/dc/elements/1.1/identifier" : [ {
    "type" : "literal",
    "value" : "A000005"
  } ],
  "http://www.library.sh.cn/ontology/noteOfSource" : [ {
    "type" : "literal",
    "value" : "中国近现代人物名号大辞典；"
  } ],
  {
    "type" : "literal",
    "value" : "20世纪中文著作者笔名录；"
  } ],
  "http://www.library.sh.cn/ontology/name" : [ {
    "type" : "uri",
    "value" : "http://data.library.sh.cn/authority/name/4gzmcio5iwp050zm"
  } ]
}
```

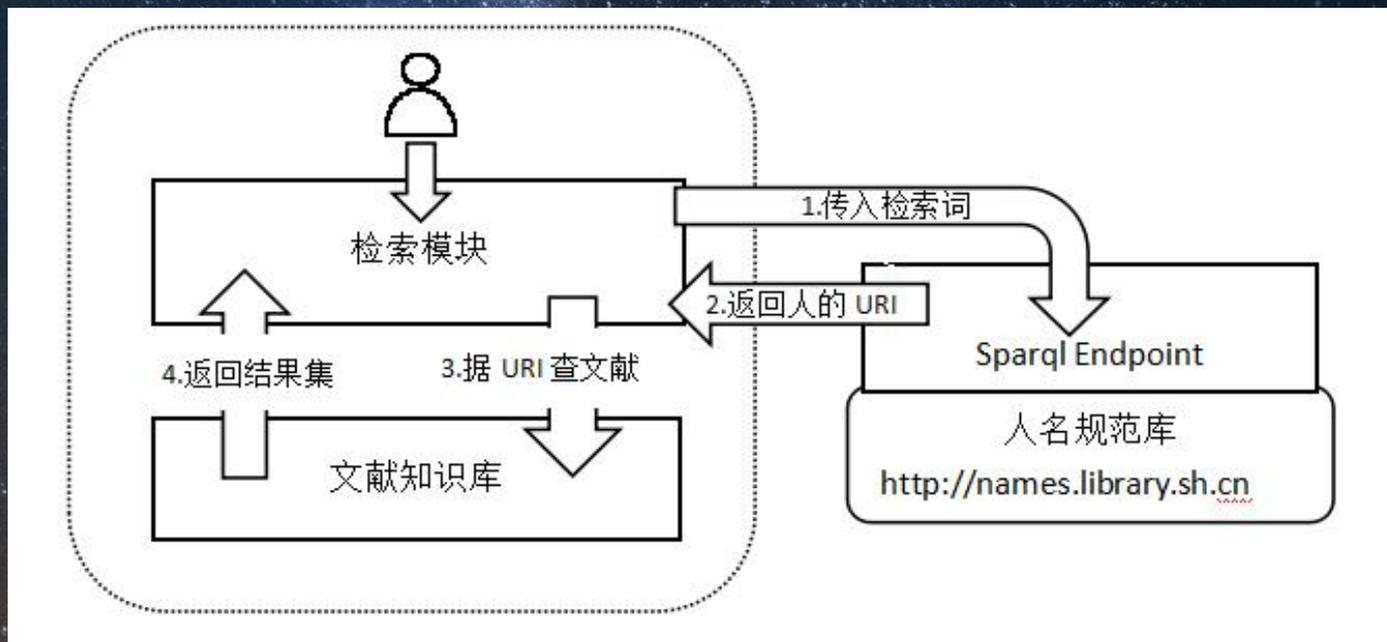
Whenever and whatever you are, only if you get the HTTP URI of a person, you can get all the RDF data of a person via the web. The data is machine readable and the format is standardized.

# 人名规范库在数字人文中的应用

The Application of Shanghai Library's Names Authority Control Knowledgebase in DH Projects

## 1. 提高查全率和查准率

Improve search precision



题名: 输入题名

作者: 茅盾

概念检索

出版年: 起始年 ~ 结束年

出版者: 输入出版者

识别码: 输入识别码

丛书: 输入丛书

查询

出版年



共匹配到 153 条记录

题名: 输入题名

作者: 茅盾

概念检索

出版年: 起始年 ~ 结束年

出版者: 输入出版者

识别码: 输入识别码

丛书: 输入丛书

查询

出版年



共匹配到 162 条记录

题名: 输入题名

作者: 沈雁冰

概念检索

出版年: 起始年 ~ 结束年

出版者: 输入出版者

识别码: 输入识别码

丛书: 输入丛书

查询

出版年



共匹配到 162 条记录

# 人名规范库在数字人文中的应用

## The Application of Shanghai Library's Names Authority Control Knowledgebase in DH Projects

### 2.动态展示实体间的关联关系

#### Display the relations among entities dynamically

盛宣怀档案知识库  
Sheng Xuanhual Archives

在此输入 输入收件人/发件人/地点/主题

盛宣怀 (字:杏生 号:止叟) 生:1844 卒:1916

1870年入李鸿章幕。1873年任上海轮船招商局会办。1880年筹办中国电报局，任总办。1892年创办北洋大学堂。其间极力协助李鸿章举办洋务。1893年筹办华盛纺织总厂，任督办，以官督商办及官商合办形式盛源洋务企业。1895年在天津创办中西学堂。1898年接办汉阳铁厂，大冶铁矿，经办芦汉铁路。1898年兼办萍乡煤矿。同年倡办中国通商银行。1900年参与英美策划的东南互保，订立《东南互保约款》。1902年任工部生侍郎。1910年任中国红十字会会长。1911年授邮传部尚书，皇族内阁成立后任邮传部大臣。与四国银行团签订币制实业借款合同，又以铁路国有名义将已归商办的川汉、粤汉干线路权作抵押，大借外债，激起铁路风潮。武昌起义后亡命日本。1916年3月25日在上海病卒。著有《愚斋存稿》、《盛宣怀未刊信稿》，刻有《常州先哲遗书》，《经世文编》、《林胡曾三公奏议》，另有《盛宣怀档案资料选编》行世。 >>>查看家谱

相关人物

- Moller
- W.K.凯梯
- 丁汝云
- 丁邦英
- 万国夔
- 严作霖
- 严信厚
- 严尔
- 严楚
- 严谦堂
- 任之骅
- 任锡汾
- 伍廷芳
- 何其垣
- 何恭寿
- 何梅
- 何炳荣
- 何德之

1844 1854 1864 1873 1874 1884 1887 1894 1904 1914

上海图书馆  
人名规范库

盛宣怀

政治家、企业家、慈善家 | 1844 - 1916

国籍: | 籍贯: 常州

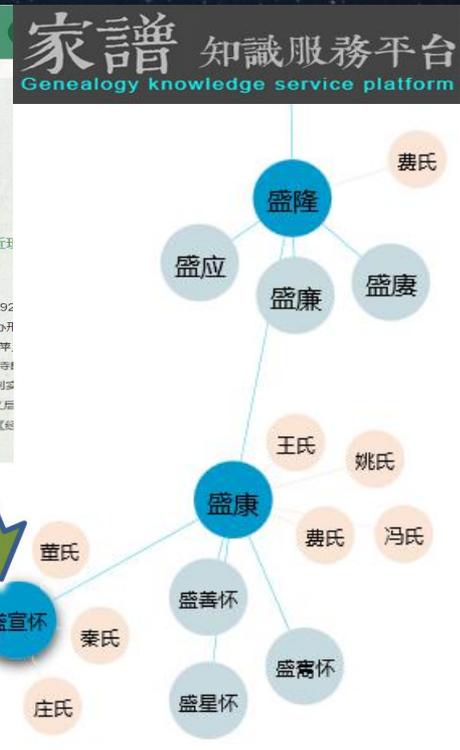
民族: 汉族 | 性别:

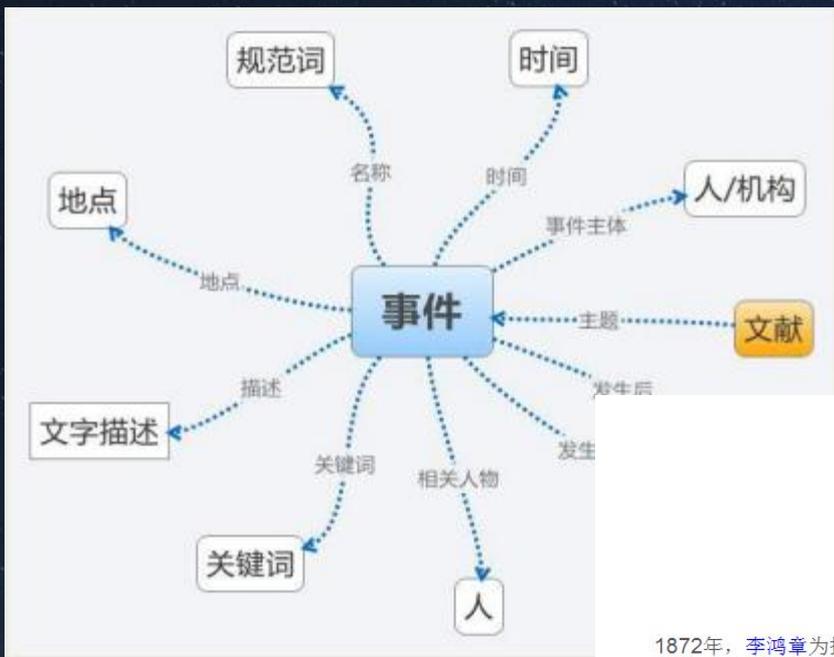
职业

上图名人手稿数据库 中国近现代人物名号大辞典 中国近现代名人辞典 中国近现代人物辞典 中国近

上图古籍数据库

1870年入李鸿章幕。1873年任上海轮船招商局会办。1880年筹办中国电报局，任总办。1892年极力协助李鸿章举办洋务。1893年筹办华盛纺织总厂，任督办，以官督商办及官商合办形式盛源洋务企业。1895年在天津创办中西学堂。1896年接办汉阳铁厂，大冶铁矿，经办芦汉铁路。1898年兼办萍乡煤矿。同年倡办中国通商银行。1900年参与英美策划的东南互保，订立《东南互保约款》。1902年任工部生侍郎。1910年任中国红十字会会长。1911年授邮传部尚书，皇族内阁成立后任邮传部大臣。与四国银行团签订币制实业借款合同，又以铁路国有名义将已归商办的川汉、粤汉干线路权作抵押，大借外债，激起铁路风潮。武昌起义后亡命日本。1916年3月25日在上海病卒。著有《愚斋存稿》、《盛宣怀未刊信稿》，刻有《常州先哲遗书》，《经世文编》、《林胡曾三公奏议》，另有《盛宣怀档案资料选编》行世。





## 盛宣怀参与创办公司大事记

1870

### 轮船招商局创办

1872年，李鸿章为挽救中国航运业，积极筹划建立新式轮船公司，以“渐分洋商之利”，在得到清廷支持后，改其为**轮船招商局**，指派专员草拟章程，在上海进行筹办。1873年1月14日，官督商办性质的**轮船招商局**正式成立，同时标志着中国近代公司制企业的出现。

#### 相关档案

轮船招商公司 轮船招商局

**[标题]** 西帮木商行事宜文

**[责任者]** 屈润培

**[主题词]** 商务 轮船招商局

**[档号]** SD059068

1/7

上一条 下一条

### 轮船招商局改名

1873年8月7日，李鸿章将**轮船招商局**从上海南永安街（今黄浦区永安路）迁至上海三马路新址，改称其**轮船招商总局**。并于同年设天津、汉口、长崎、香港等19个分局。

#### 相关档案

轮船招商总局 轮船招商局

**[标题]** 李鸿章礼盛宣怀文

**[责任者]** 李鸿章

**[主题词]** 任官 借贷 制钱 唐廷枢 天津练谕局 朱其昂 漕粮 盛宣怀 轮船招商局 运输 采办

**[档号]** SD059551

1/19

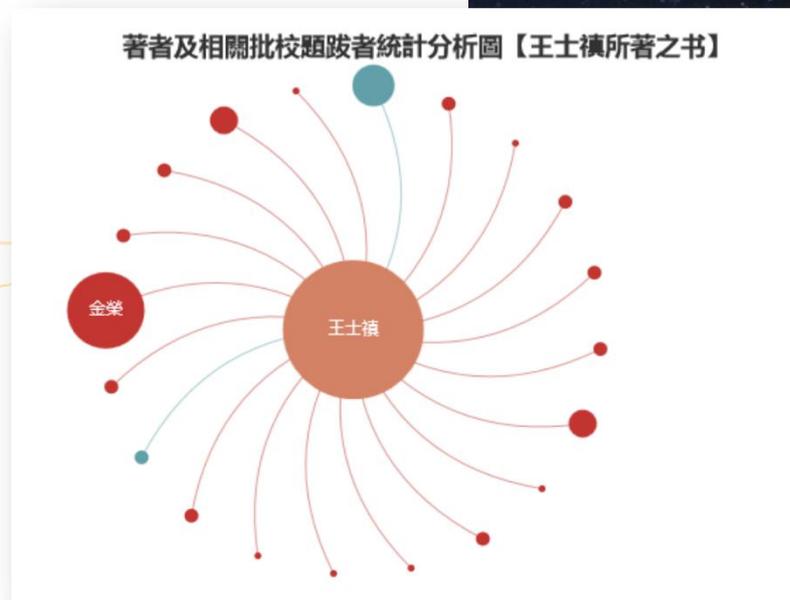
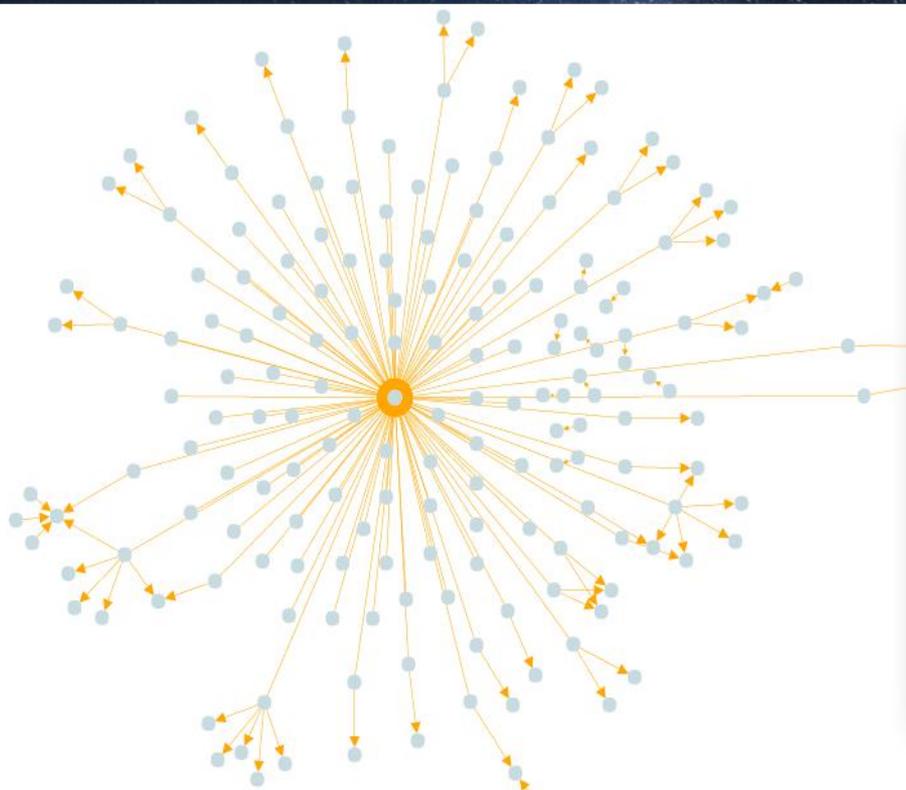
上一条 下一条

# 人名规范库在数字人文中的应用

The Application of Shanghai Library's Names Authority Control Knowledgebase in DH Projects

## 3.助力社会网络关系分析

Help analysis of social network relationship



# 人名规范库在数字人文中的应用

## The Application of Shanghai Library's Names Authority Control Knowledgebase in DH Projects

### 4.支持命名实体识别Support Name Entity Recognition

上海图书馆开发数据平台  
数据采集转换工具

任务

- 百科任务
- 采集数据
- 数据加工
- 数据转换
- Sparql任务
- 采集数据
- 数据加工
- 数据转换
- 命名实体识别

模板

- 过滤模板
- 抽取模板
- RDF映射模板

资源

- Iri列表

工具

- 实体识别

实体识别 多条识别

识别结果

顾廷龙 应 叶景葵 张元济 之道，返沪创办 私立合众图书馆，历任 暨南大学、 光华大学 教授。顾廷龙 (1904~1998)，现代图书馆事业家、古籍版本目录学家、古文字学家、书法家。字起潜，别号菊影。江苏苏州人，世居上海。幼承家学，喜习诗文书法。读大学时从 胡朴安、 闻宥、 刘三 等习古文字与书法。1931年入 燕京大学 当研究生，专研古文字学，1932年毕业获文学学士 和 文学硕士学位，留校在图书馆工作，业余 兼 与 容庚、 商承祚 等人研究书法。新中国建立后，历任 上海市历史文献图书馆 馆长， 上海图书馆 馆长、研究员， 《辞海》 编委、图书馆分科主编，文化部国家文物鉴定委员会委员，国务院古籍整理出版规划小组顾问，中国图书馆学会第一、二、三届副理事长， 华东师范大学、 复旦大学 兼职教授， 上海市 第三至五届人民代表大会代表， 上海市 政协第四至六届常务委员，中国书法家协会名誉理事。幼时随父 顾元昌 学 中名书家 习书法，临摹 吴大澂 《论语》《李经》 篆书，由篆隶上溯古籀。擅多种书体。尤 精 精全文大篆，以丰茂雄深、质朴古雅见长。隶书喜临《石门颂》。楷书始习欧阳询，领略“刚健险劲，法度森严”的风格；继学 赵孟頫，寻求“意在笔先，笔到法随”；再学 苏东坡，更显笔力雄健，姿态自如。作品多次在国内外重要书展展出，并为多处博物馆、纪念馆收藏。1996年 上海图书馆 新馆落成时，举办了“顾廷龙 书法展”，同时出版《顾廷龙 书法展》，同时出版《顾廷龙 书法选集》。1963年 和 1979年为 中国书法家代表团 和 上海 书法家代表团成员两度访问日本。1978年参加中日书法讨论会及兰亭书会。主编《中国丛书综录》、《中国古籍善本书目》、《续修四库全书》等，1991年 上海市 政府授奖记大功，1995年获文化部褒奖。出版有《吴意斋先生年谱》、《古匋文彙录》、《董氏四当斋书目》、《明代版本图景初探》(与潘承弼合编)、《尚书文字合编》(与顾颉刚合编)等。

文本提交

人名 姓氏 地名 机构

提交

实体详情

顾廷龙	
顾廷龙	
相关家谱文 献	<a href="http://data.library.sh.cn/jp/resource/work/773dm3gea34ejftm">http://data.library.sh.cn/jp/resource/work/773dm3gea34ejftm</a>
角色	<a href="http://data.library.sh.cn/jp/vocab/ancestor/xian-zu">http://data.library.sh.cn/jp/vocab/ancestor/xian-zu</a>
姓氏	<a href="http://data.library.sh.cn/authority/familyname/wekosyv124e46pi3">http://data.library.sh.cn/authority/familyname/wekosyv124e46pi3</a>



谢谢THANKS!

[xtykc@yeah.net](mailto:xtykc@yeah.net)