Dublin Core™ Metadata Initiative

The Integrated Workfolw of Entity Management and Service for Digital Humanities

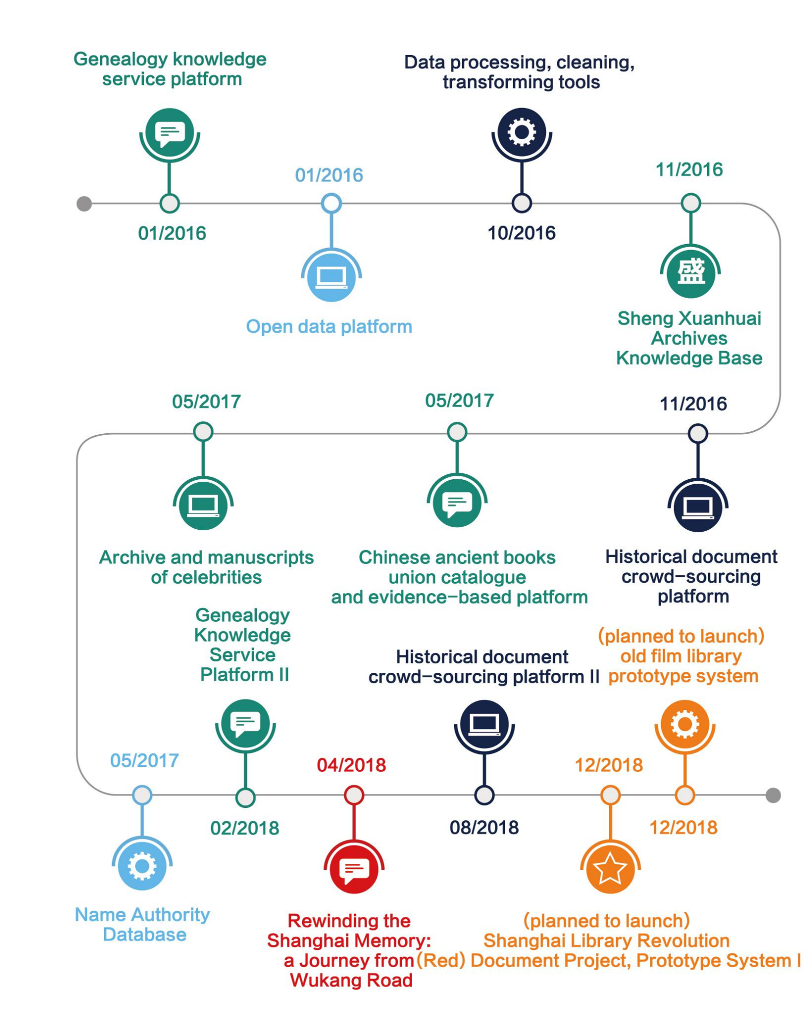
Cuijuan Xia   Shanghai Library   Oct. 14,2021

DCMI 2021 VIRTUAL
Online, 2021 October 4-15

01 Background

Current Situation and Problems

From Digital Library systems to semantic knowledgebases



01/2016
Genealogy knowledge service platform

01/2016
Open data platform

10/2016
Data processing, cleaning, transforming tools

11/2016
Sheng Xuanhuai Archives Knowledge Base

11/2016
Historical document crowd−sourcing platform

05/2017
Chinese ancient books union catalogue and evidence−based platform

05/2017
Archive and manuscripts of celebrities

05/2017
Name Authority Database

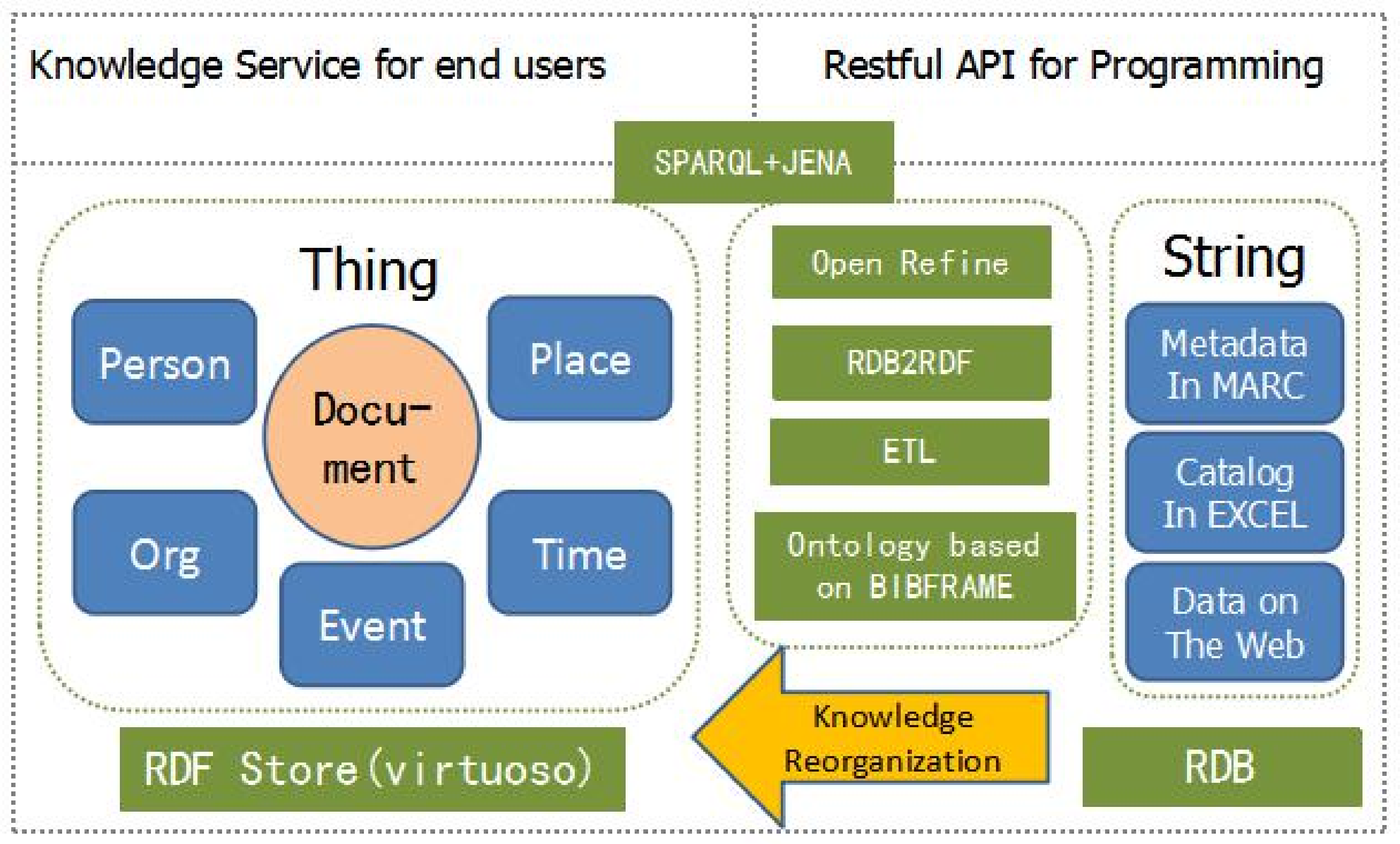
02/2018
Genealogy Knowledge Service Platform II

04/2018
Rewinding the Shanghai Memory: a Journey from Wukang Road

08/2018
Historical document crowd−sourcing platform II

12/2018
(planned to launch) Shanghai Library Revolution (Red) Document Project, Prototype System I

12/2018
(planned to launch) old film library prototype system



Knowledge Service for end users

Restful API for Programming

SPARQL+JENA

Thing

Person

Place

Docu-ment

Org

Time

Event

Open Refine

RDB2RDF

ETL

Ontology based on BIBFRAME

String

Metadata In MARC

Catalog In EXCEL

Data on The Web

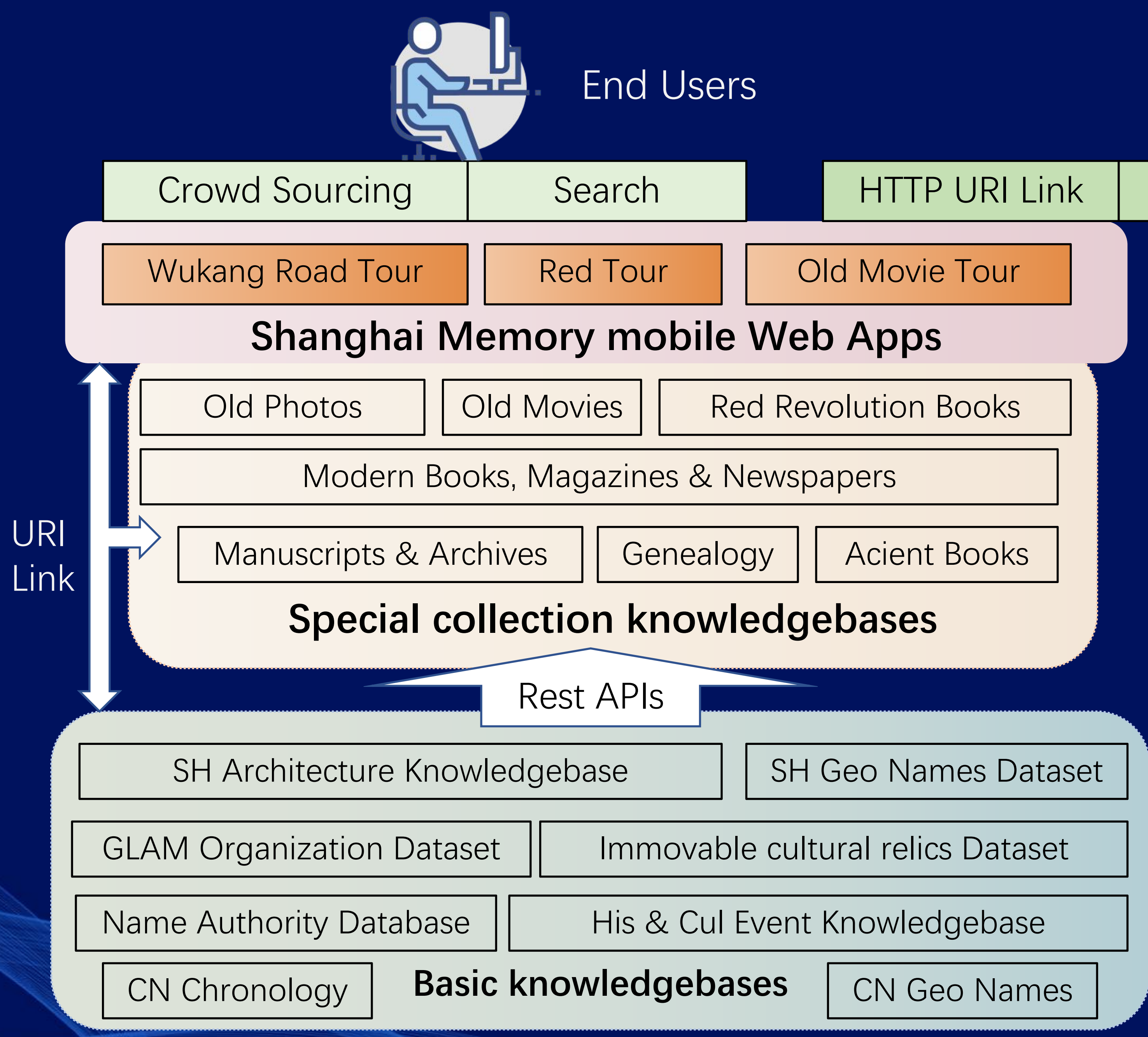
RDF Store(virtuoso)

Knowledge Reorganization

RDB

Process

- Transform all metadata records of different resources in different formats into RDF triples. Take authors and contributors, publishers and organizations, events and places as entities but not as strings.
- Give every entity cool URI as global identifier and locator. And then link them together after Name Entity Recognization(NER) and entity disambiguation.
- Enrich more semantic data about the entities by extracting structured data from the content of digital resource objects or open datasets on the Web like wiki data.
- Provide authority control and knowledge linking service among resources of different kinds of collections in a web-scale. Provide digital Humanities services for researchers, and open data APIs for the third party developers.



End Users

Developers

Crowd Sourcing | Search

HTTP URI Link | Restful API | Sparql Endpoint

Wukang Road Tour | Red Tour | Old Movie Tour

Shanghai Memory mobile Web Apps

URI Link

Old Photos | Old Movies | Red Revolution Books

Modern Books, Magazines & Newspapers

Manuscripts & Archives | Genealogy | Acient Books

Special collection knowledgebases

Rest APIs

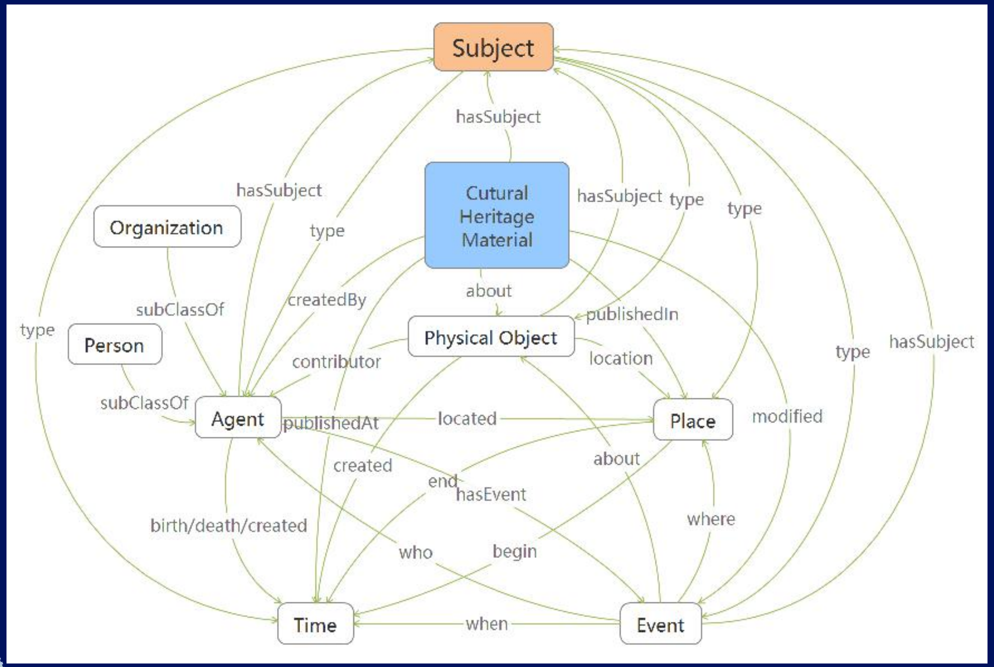
SH Architecture Knowledgebase | SH Geo Names Dataset

GLAM Organization Dataset | Immovable cultural relics Dataset

Name Authority Database | His & Cul Event Knowledgebase

CN Chronology | CN Geo Names

Basic knowledgebases



Linked entities and data of different knowledgebases with URIs semanticly by the One Ontology Abstract Model.

Difficulties and Problems

- Data Cleaning and NER.
    - Have to correct the Inconsistency of metadata records in digital library systems and deal with the entity disambiguation manually.

- Decentralized process of entities extraction, creation, modification, publishing, and management.

Solutions

- **Integrated Workflow**
  - Bridge the gaps among the process of entities management(including extraction, creation, modification, publishing, interlinking) and Service(federal search, visualizaiton).
- **New Technology application**
  - Machine Learning to make the OCR and NER work more efficiently.
- **One platform**
  - support the **Integrated Workflow and new technology application.**
  - support the services(across knowledgebases and visulizations for SNS and spacial-temporal analysis) based on entities and the relations among entites.

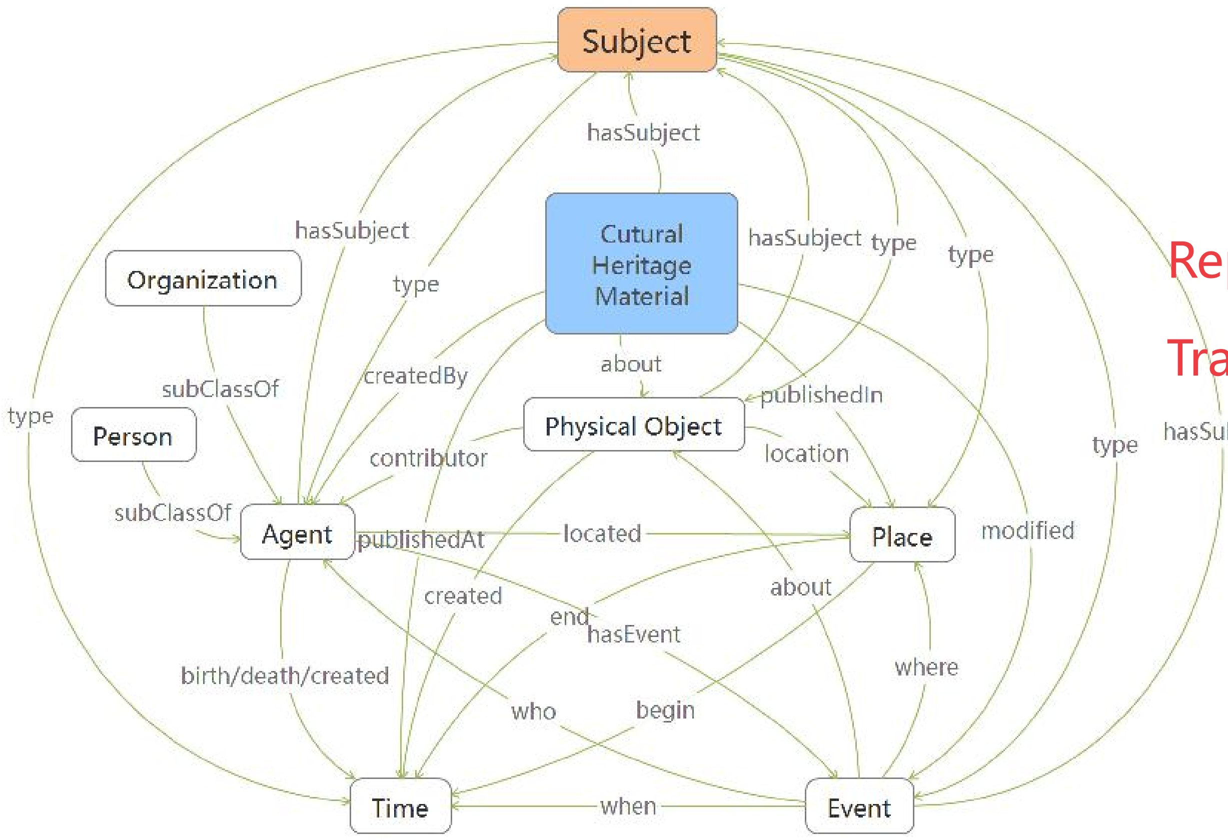
02

Semantic Content Architecture

The Semantic Content Architecture of Entity Management

Ontology Abstract Model of All Entities

http://data.library.sh.cn/ont/ontology/search

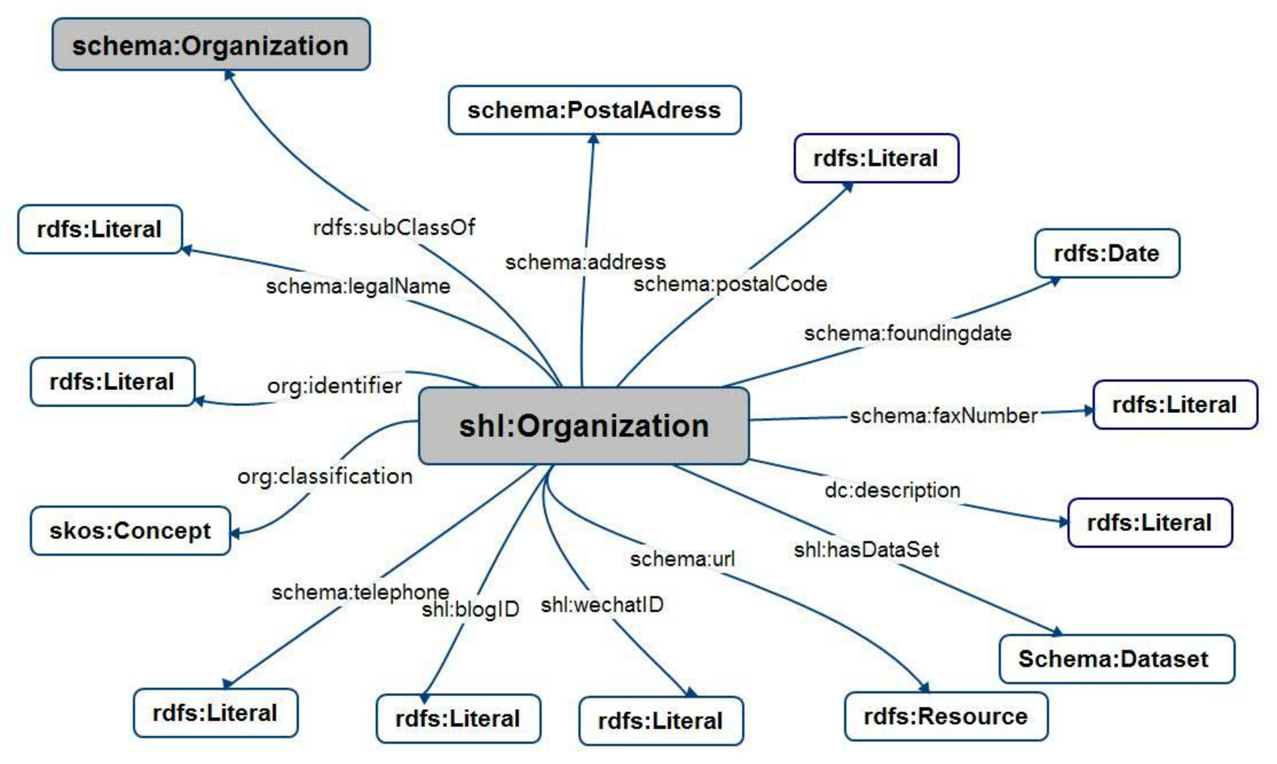


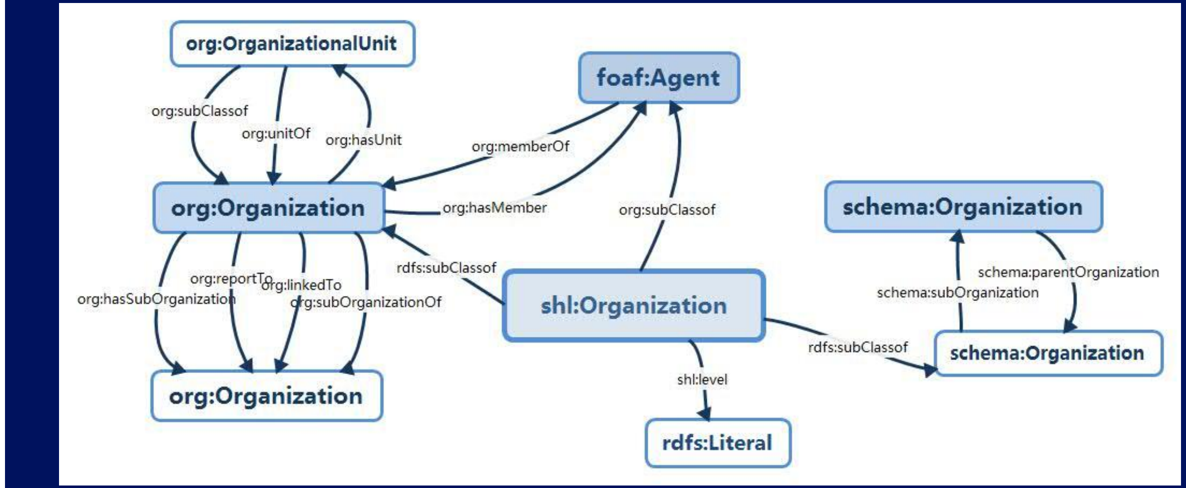
Modeling	Ontology	
Representation	RDF	
Transformation	XML Turtle N3 JSON	
Storage	Triple Store	
Search	SPARQL	
Publish	HTTP	
Access	Cool URI	Restful API

one unified ontology for knowledge modeling, RDF sepecification for consistent knowledge representation

夏翠娟. 文化记忆资源的知识融通：从异构资源元数据应用纲要到一体化本体设计.图书情报知识,2021(1)

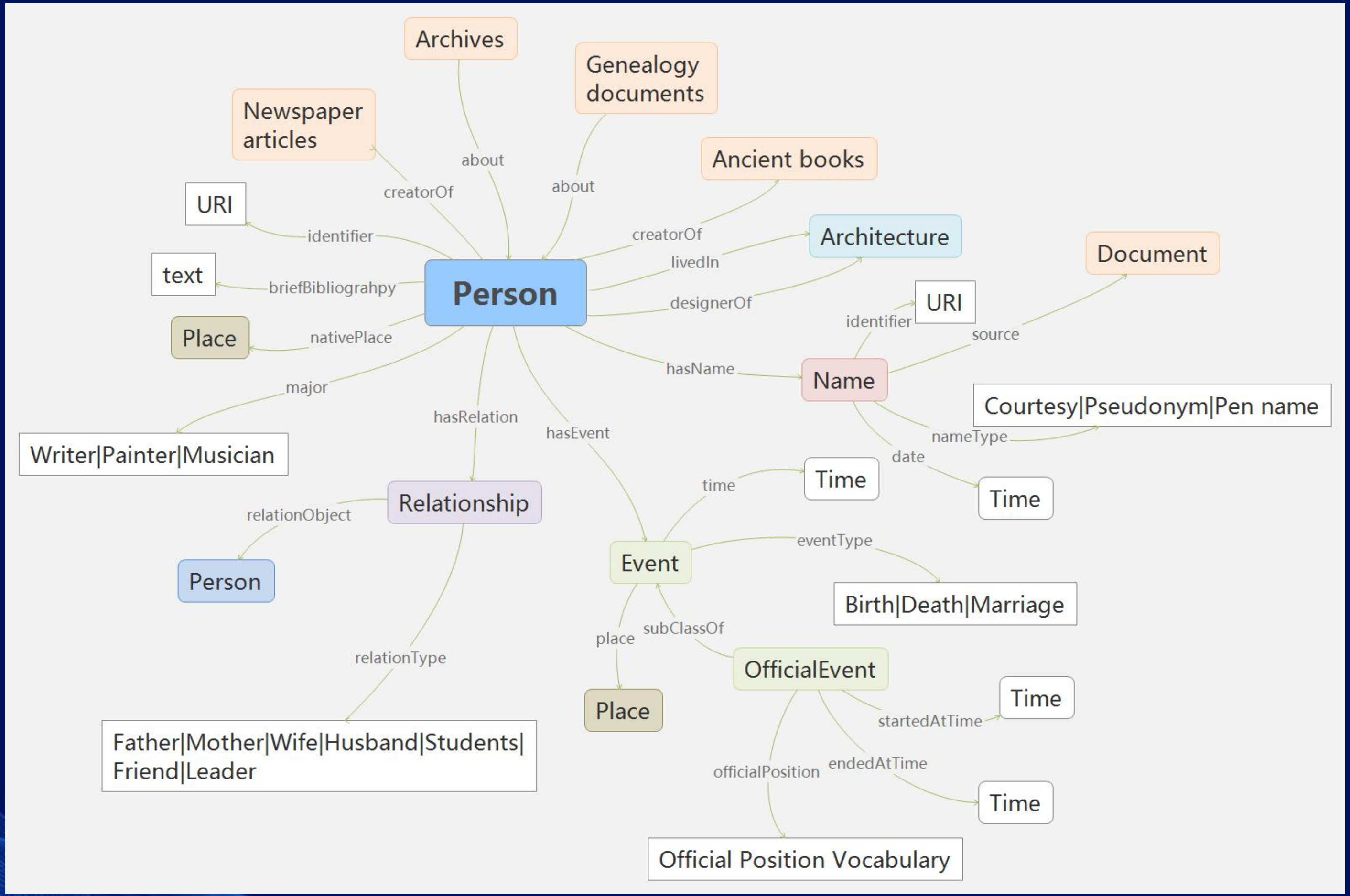
Ontology Application Profile of Organization

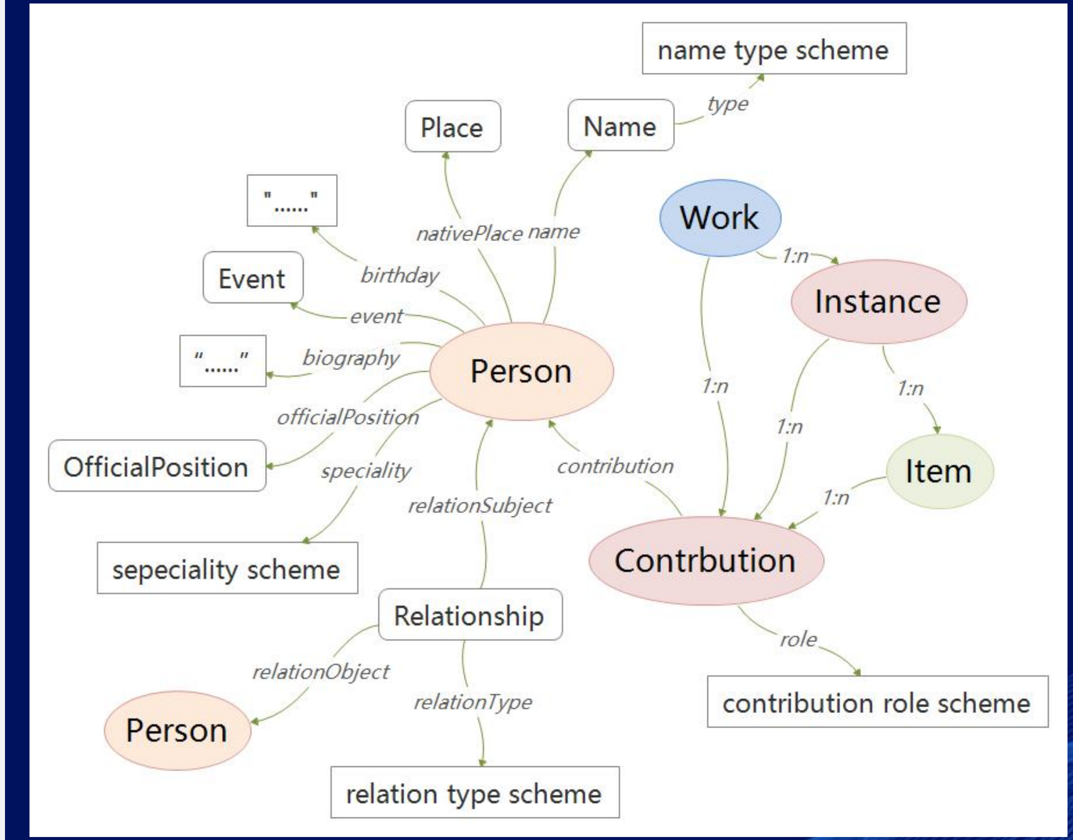




金家琴,夏翠娟.数字人文数据基础设施建设中机构本体的构建：研究和应用[J].图书馆论坛,2020,40(04):30-39.

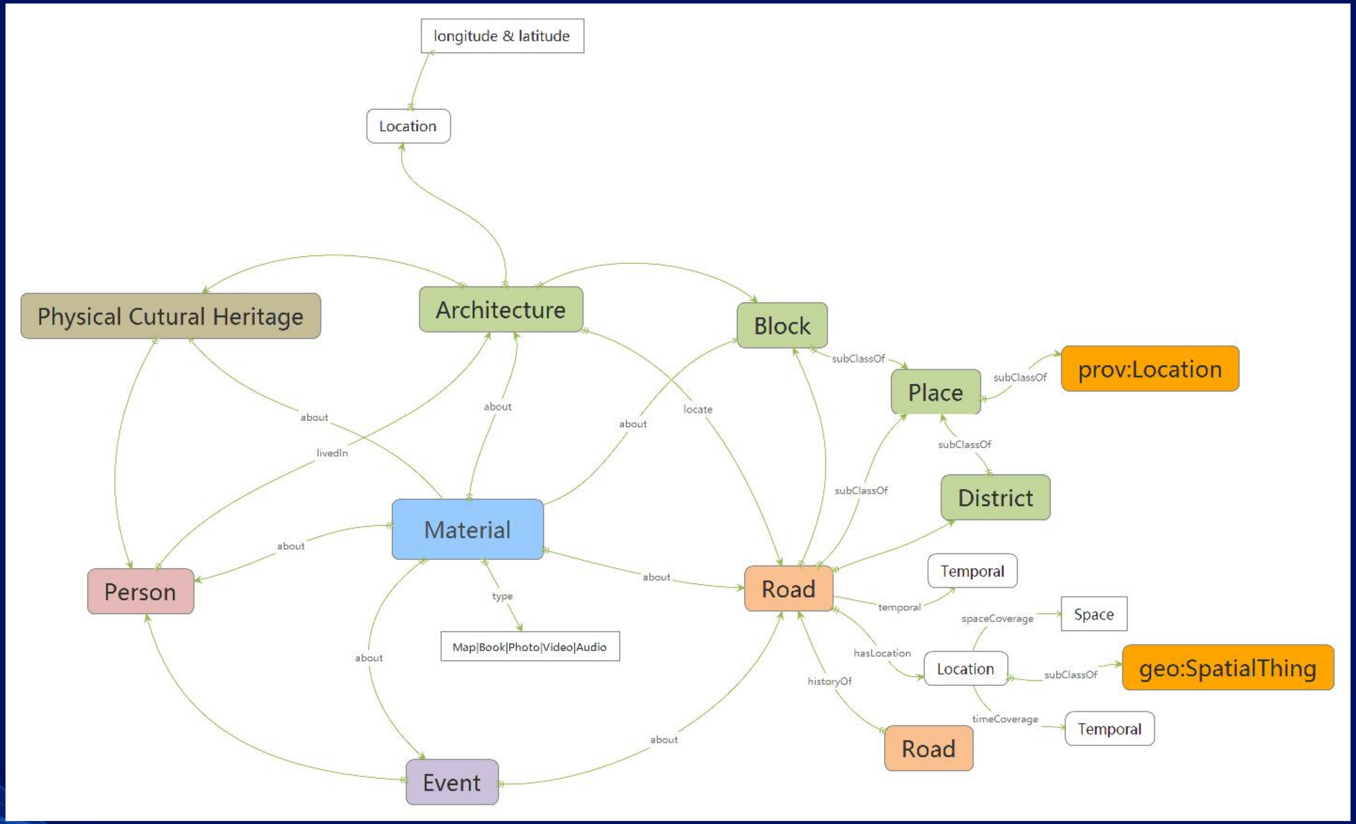
Ontology Application Profile of Person

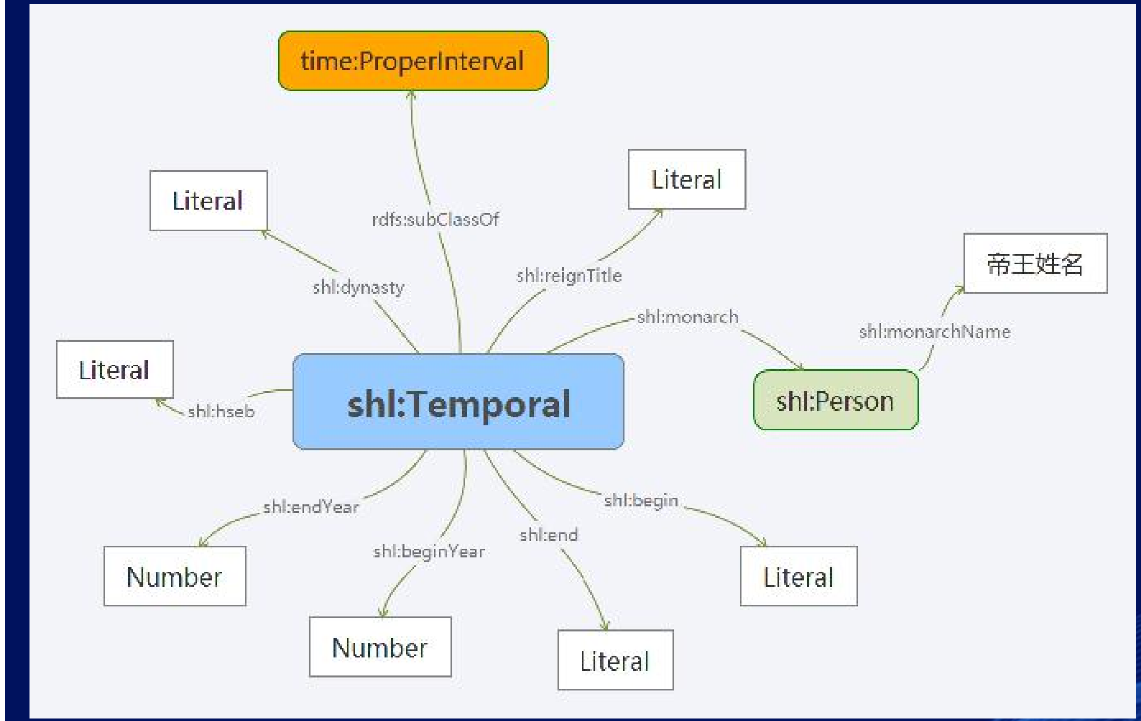




Xia, Cuijuan, and Liu, Wei. "Name Authority Control in Digital Humanities: Building a Name Authority Database of Shanghai Library." International Journal of Libraryship,3.1(2018):21.

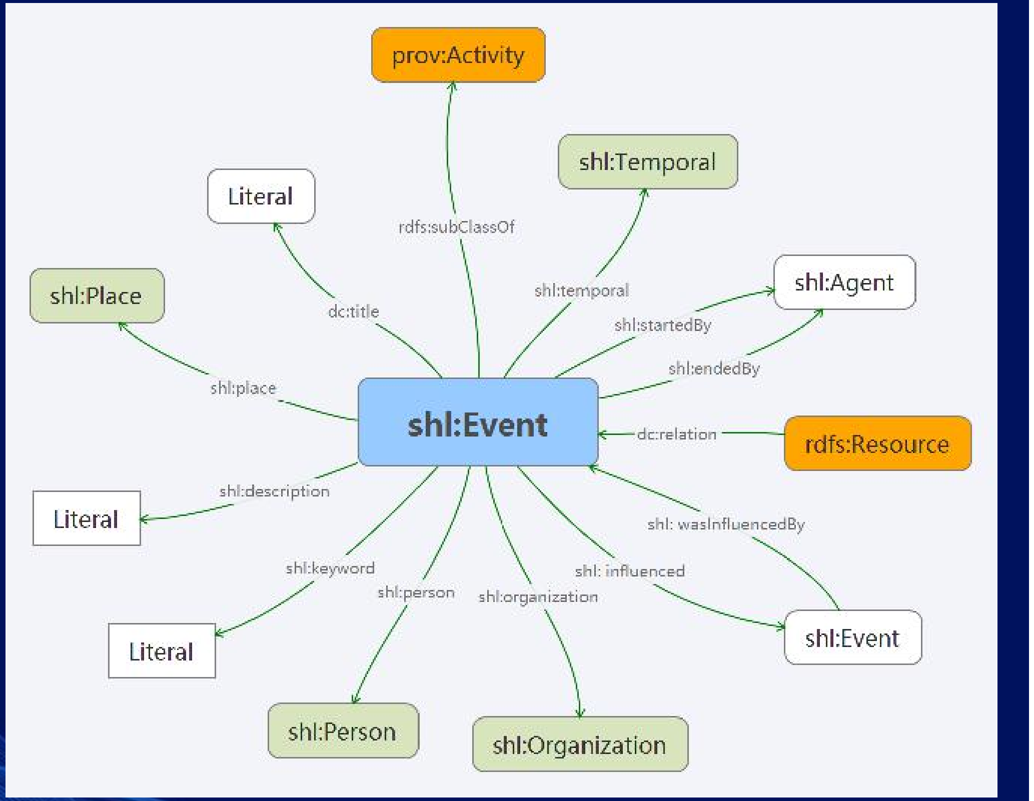
Ontology Application Profile of Place and Temporal

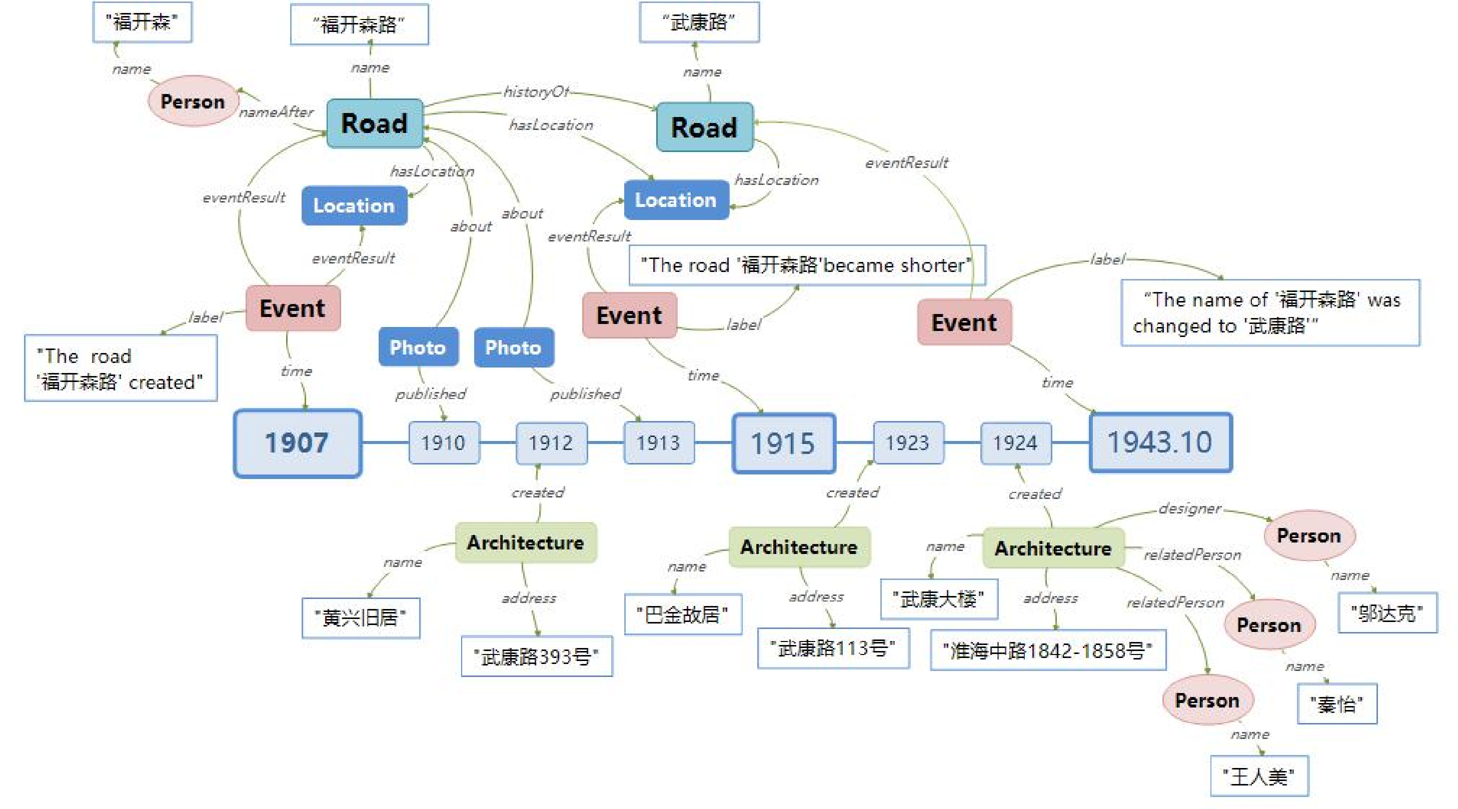




Xia Cuijuan,Wang Lihua,Liu Wei. Shanghai memory as a digital humanities platform to rebuild the history of the city. Digital Scholarship in the Humanities, 2021(3) https://academic.oup.com/dsh/advance-article/doi/10.1093/llc/fqab023/6178577?guestAccessKey=2a5451a0-5080-4b49-9e5f-bf0fae4f2e10

Ontology Application Profile of Event





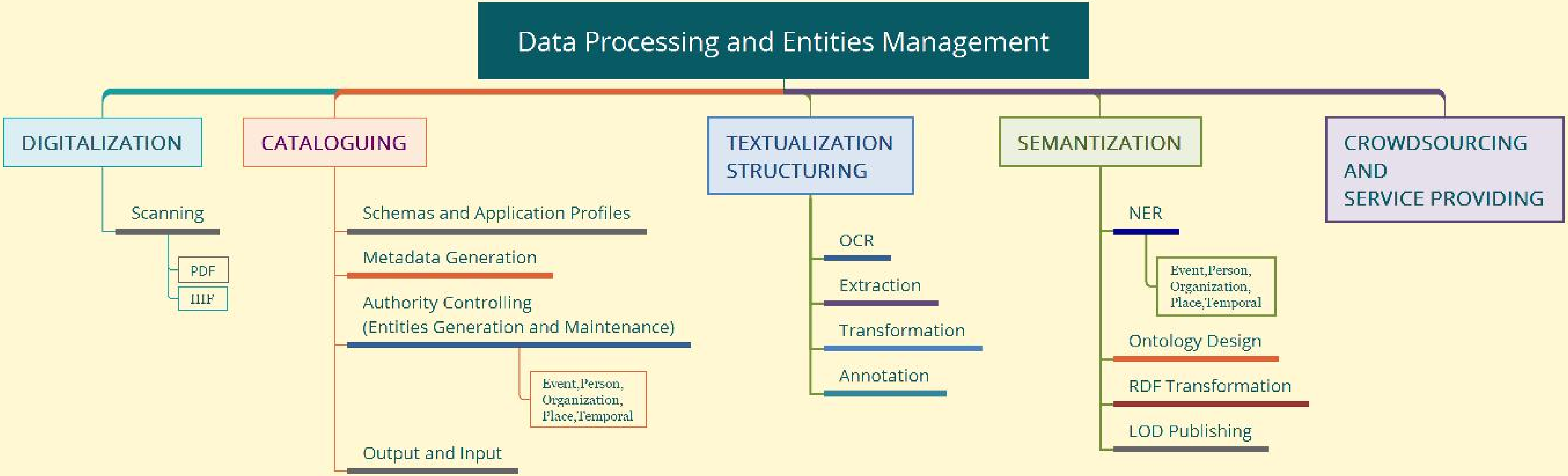
Xia Cuijuan,Wang Lihua,Liu Wei. Shanghai memory as a digital humanities platform to rebuild the history of the city. Digital Scholarship in the Humanities, 2021(3) https://academic.oup.com/dsh/advance-article/doi/10.1093/llc/fqab023/6178577?guestAccessKey=2a5451a0-5080-4b49-9e5f-bf0fae4f2e10

03

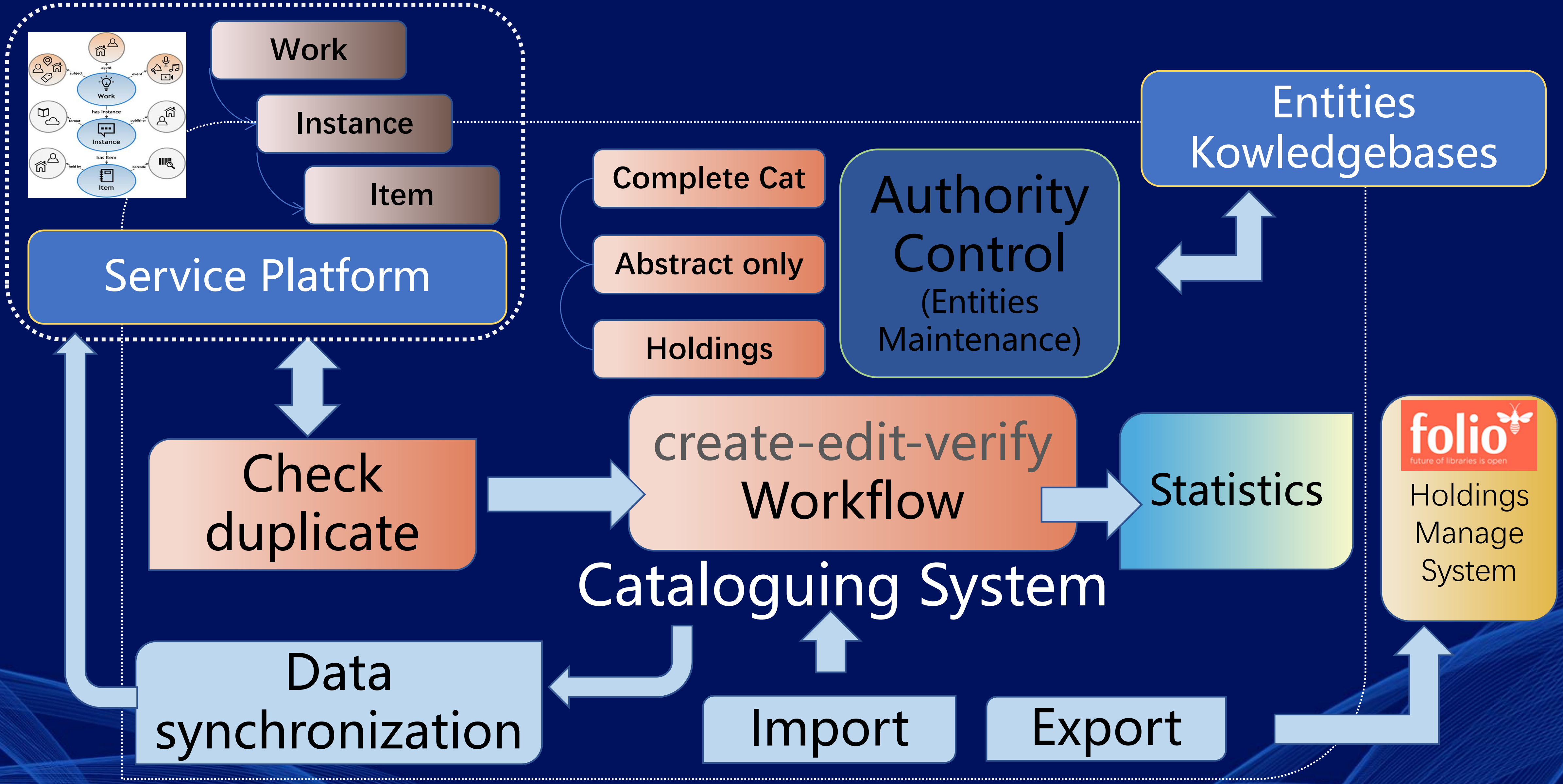
Integrated Workflow

Integrated workflow to bridge the gaps among the different stages during the data processing

Integrated Workflow for
Data Processing and Entity Management



Genealogy Union Cataloging System as an Example



Work

Instance

Item

Service Platform

Complete Cat

Abstract only

Holdings

Authority Control (Entities Maintenance)

Entities Kowledgebases

Check duplicate

create-edit-verify Workflow

Statistics

Holdings Manage System

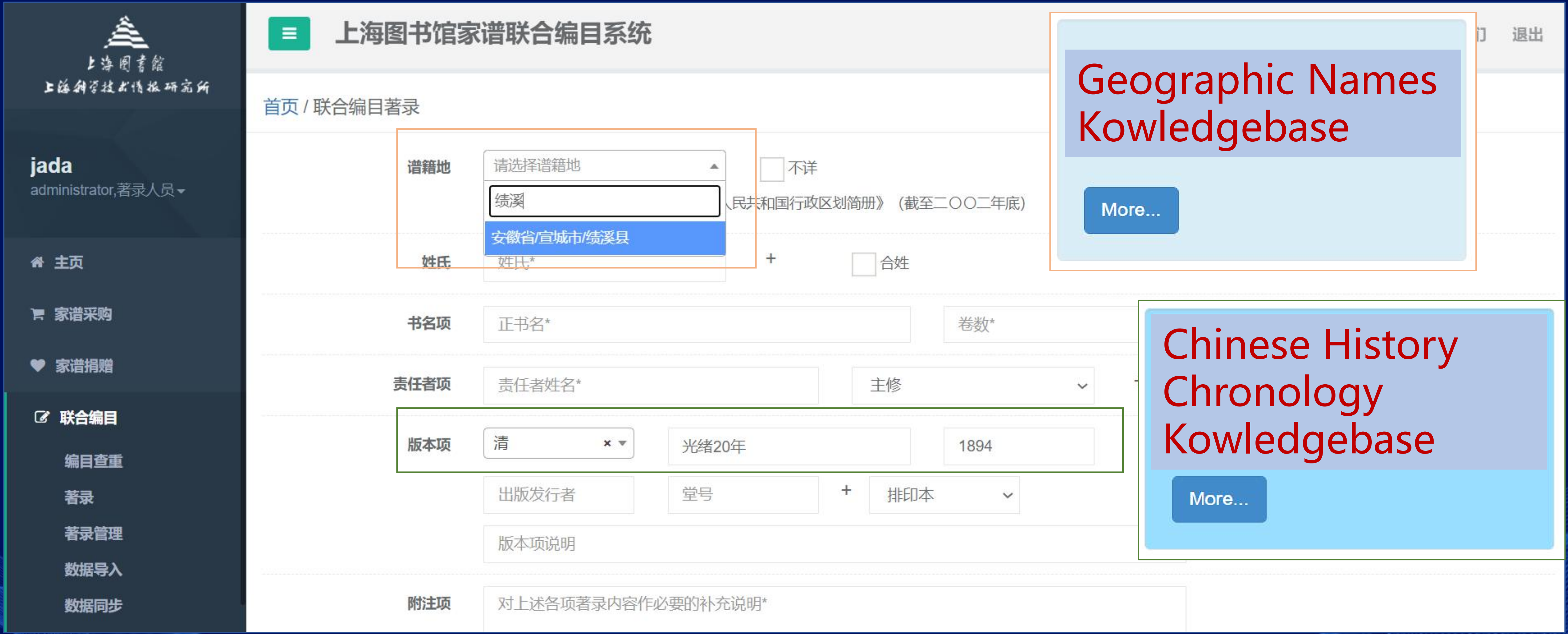
Cataloguing System

Data synchronization

Import

Export

Authority Control (Entities Maintenance) during cataloging



上海图书馆
上海科学技术情报研究所

jada
administrator,著录人员

- 主页
- 家谱采购
- 家谱捐赠
- 联合编目
  - 编目查重
  - 著录
  - 著录管理
  - 数据导入
  - 数据同步

上海图书馆家谱联合编目系统

退出

首页 / 联合编目著录

谱籍地：请选择谱籍地 □ 不详

绩溪

安徽省/宣城市/绩溪县

民共和国行政区划简册》（截至二〇〇二年底）

姓氏：姓氏* + □ 合姓

书名项：正书名* 卷数*

责任者项：责任者姓名* 主修

版本项：清 光绪20年 1894

出版发行者 堂号 + 排印本

版本项说明

附注项：对上述各项著录内容作必要的补充说明*

Geographic Names Kowledgebase

More...

Chinese History Chronology Kowledgebase

More...

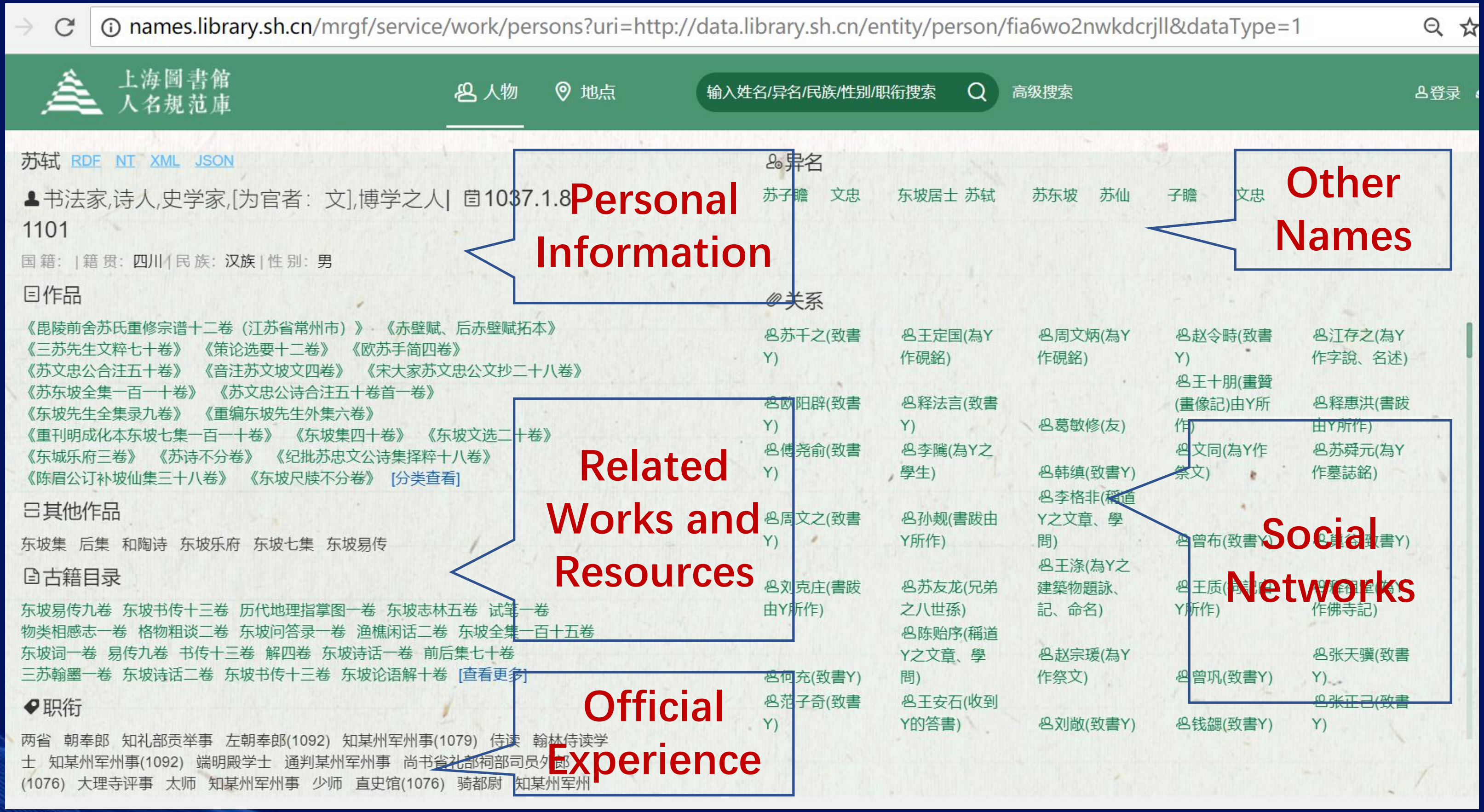
https://jplb.library.sh.cn

04

Services Support

SNS and spacial-temporal analysis

Name Authority Database as a Data Linking Hub



names.library.sh.cn/mrgf/service/work/persons?uri=http://data.library.sh.cn/entity/person/fia6wo2nwkdcrjll&dataType=1

苏轼 RDF NT XML JSON

书法家,诗人,史学家,[为官者：文],博学之人| 1037.1.8 1101

国 籍：| 籍 贯：四川 | 民 族：汉族 | 性 别：男

作品

《昆陵前舍苏氏重修宗谱十二卷（江苏省常州市）》 《赤壁赋、后赤壁赋拓本》 《三苏先生文粹七十卷》 《策论选要十二卷》 《欧苏手简四卷》 《苏文忠公合注五十卷》 《音注苏文坡文四卷》 《宋大家苏文忠公文抄二十八卷》 《苏东坡全集一百一十卷》 《苏文忠公诗合注五十卷首一卷》 《东坡先生全集录九卷》 《重编东坡先生外集六卷》 《重刊明成化本东坡七集一百一十卷》 《东坡集四十卷》 《东坡文选二十卷》 《东城乐府三卷》 《苏诗不分卷》 《纪批苏忠文公诗集择粹十八卷》 《陈眉公订补坡仙集三十八卷》 《东坡尺牍不分卷》 [分类查看]

其他作品

东坡集 后集 和陶诗 东坡乐府 东坡七集 东坡易传

古籍目录

东坡易传九卷 东坡书传十三卷 历代地理指掌图一卷 东坡志林五卷 试笔一卷 物类相感志一卷 格物粗谈二卷 东坡问答录一卷 渔樵闲话二卷 东坡全集一百十五卷 东坡词一卷 易传九卷 书传十三卷 解四卷 东坡诗话一卷 前后集七十卷 三苏翰墨一卷 东坡诗话二卷 东坡书传十三卷 东坡论语解十卷 [查看更多]

职衔

两省 朝奉郎 知礼部贡举事 左朝奉郎(1092) 知某州军州事(1079) 侍读 翰林侍读学士 知某州军州事(1092) 端明殿学士 通判某州军州事 尚书省礼部祠部司员外郎(1076) 大理寺评事 太师 知某州军州事 少师 直史馆(1076) 骑都尉 知某州军州

异名

苏子瞻 文忠 东坡居士 苏轼 苏东坡 苏仙 子瞻 文忠

关系

苏千之(致書Y) 王定国(為Y作硯銘) 周文炳(為Y作硯銘) 赵令畤(致書Y) 江存之(為Y作字說、名述)
王十朋(畫贊(畫像記)由Y所作)
欧阳辟(致書Y) 释法言(致書Y) 葛敏修(友) 释惠洪(書跋由Y所作)
傅尧俞(致書Y) 李隲(為Y之學生) 韩缜(致書Y) 文同(為Y作祭文) 苏舜元(為Y作墓誌銘)
李格非(稱道Y之文章、學問)
周文之(致書Y) 孙觌(書跋由Y所作) 曾布(致書Y) [illegible](致書Y)
王涤(為Y之建築物題詠、記、命名)
刘克庄(書跋由Y所作) 苏友龙(兄弟之八世孫) 王质([illegible]由Y所作) [illegible]作佛寺記)
陈贻序(稱道Y之文章、學問)
何充(致書Y) 赵宗瑗(為Y作祭文) 曾巩(致書Y) 张天骥(致書Y)
范子奇(致書Y) 王安石(收到Y的答書) 刘敞(致書Y) 钱勰(致書Y) 张正己(致書Y)

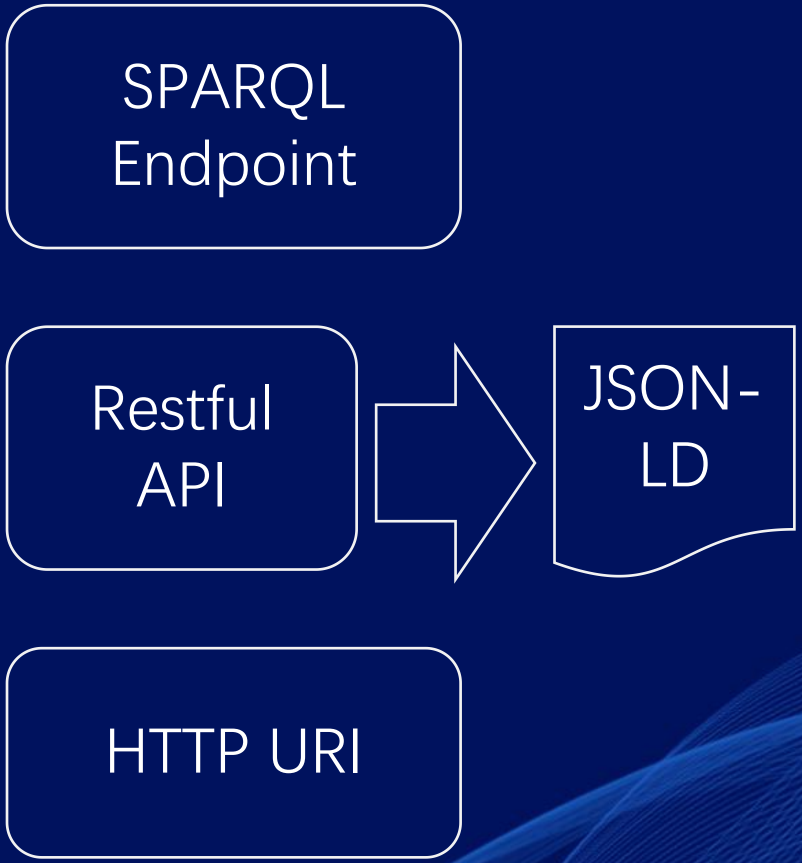
Personal Information

Other Names

Related Works and Resources

Official Experience

Social Networks



SPARQL Endpoint

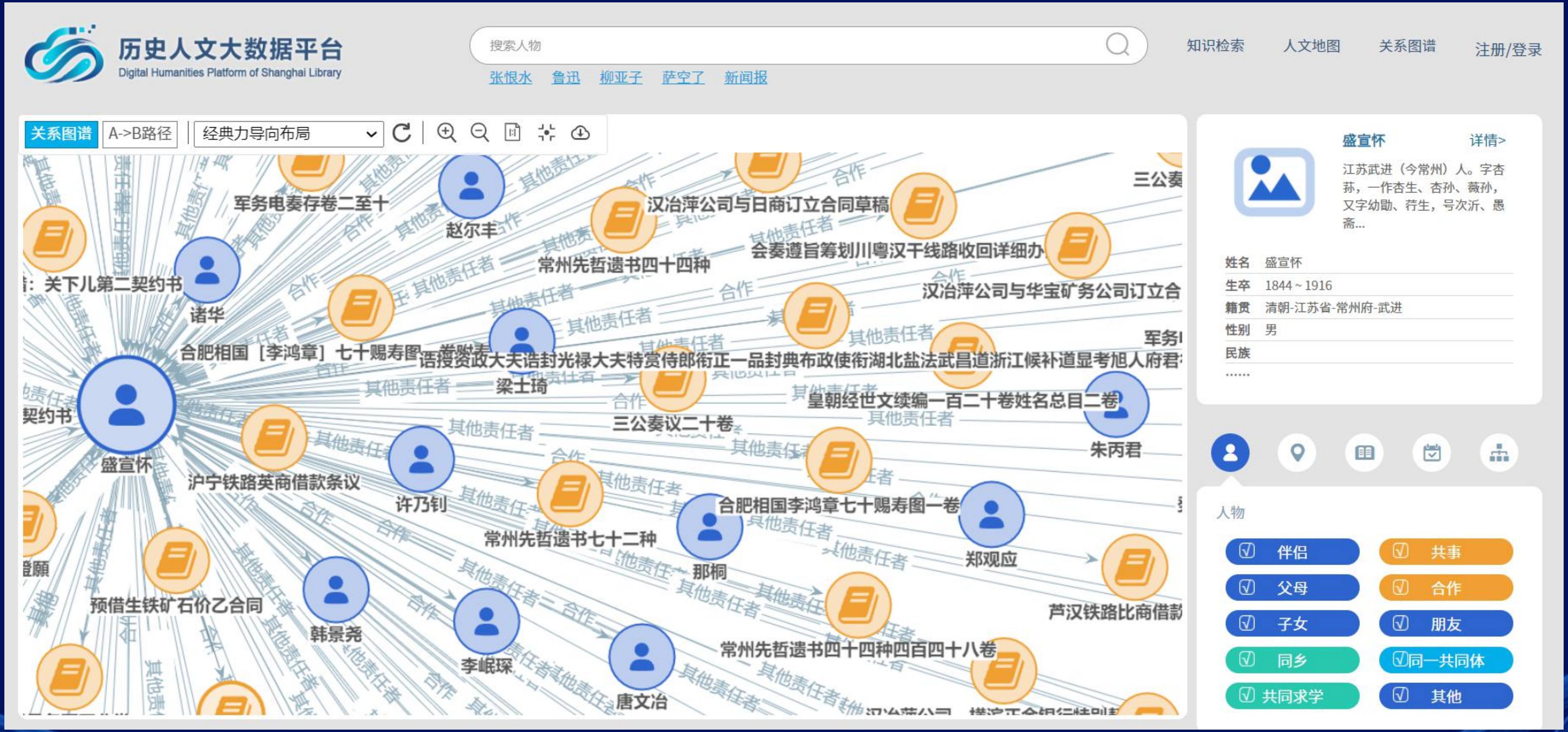
Restful API

JSON-LD

HTTP URI

http://names.library.sh.cn

SNS Social Network analysis



历史人文大数据平台
Digital Humanities Platform of Shanghai Library

搜索人物

张恨水 鲁迅 柳亚子 萨空了 新闻报

知识检索 人文地图 关系图谱 注册/登录

关系图谱 A->B路径 经典力导向布局

盛宣怀
军务电奏存卷二至十
赵尔丰
汉冶萍公司与日商订立合同草稿
三公奏
诸华
常州先哲遗书四十四种
会奏遵旨筹划川粤汉干线路收回详细办
汉冶萍公司与华宝矿务公司订立合
关下儿第二契约书
合肥相国［李鸿章］七十赐寿图
诰授资政大夫诰封光禄大夫特赏侍郎衔正一品封典布政使衔湖北盐法武昌道浙江候补道显考旭人府君
梁士琦
契约书
三公奏议二十卷
皇朝经世文续编一百二十卷姓名总目二卷
朱丙君
沪宁铁路英商借款条议
许乃钊
常州先哲遗书七十二种
合肥相国李鸿章七十赐寿图一卷
那桐
郑观应
预借生铁矿石价乙合同
韩景尧
李岷琛
唐文治
芦汉铁路比商借款
常州先哲遗书四十四种四百四十八卷
其他责任者
合作

盛宣怀 详情>

江苏武进（今常州）人。字杏荪，一作杏生、杏孙、薇孙，又字幼勖、荇生，号次沂、愚斋...

姓名	盛宣怀
生卒	1844 ~ 1916
籍贯	清朝-江苏省-常州府-武进
性别	男
民族	

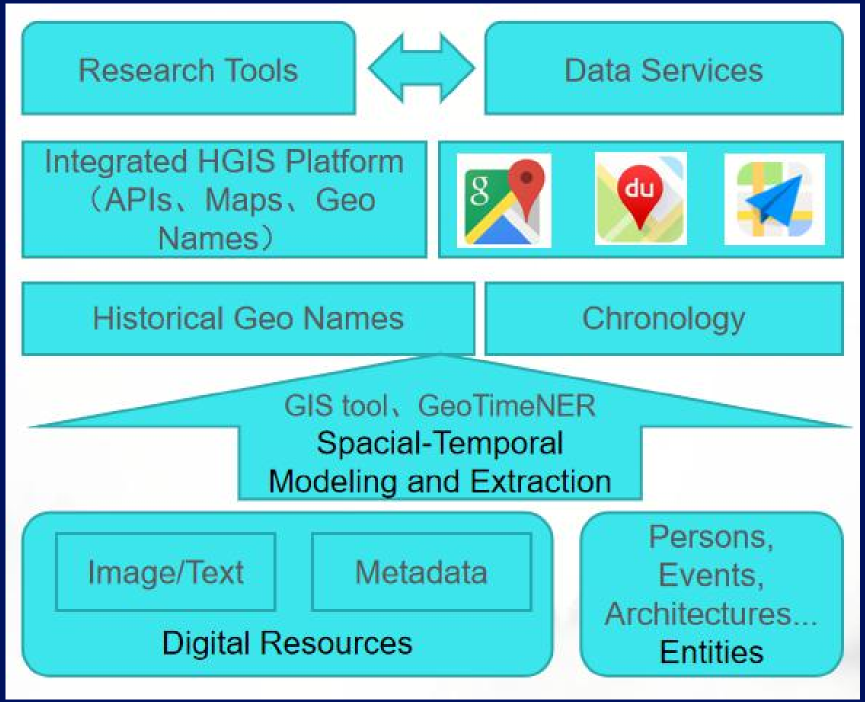
......

人物

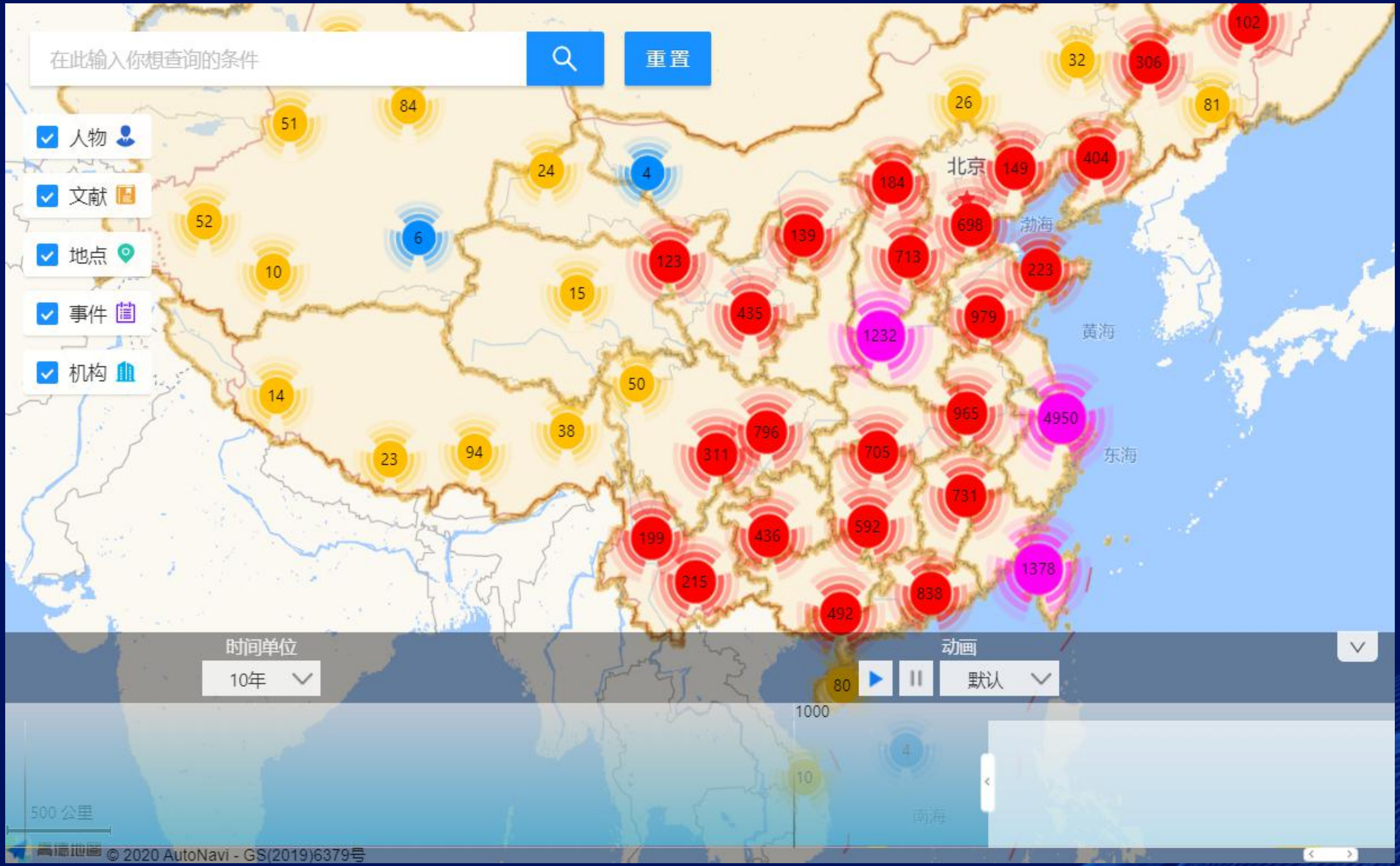
☑ 伴侣 ☑ 共事
☑ 父母 ☑ 合作
☑ 子女 ☑ 朋友
☑ 同乡 ☑同一共同体
☑ 共同求学 ☑ 其他

http://dhc.library.sh.c

HGIS Spatial-temporal analysis


Research Tools
Data Services
Integrated HGIS Platform (APIs、Maps、Geo Names)
du
Historical Geo Names
Chronology
GIS tool、GeoTimeNER
Spacial-Temporal Modeling and Extraction
Image/Text
Metadata
Digital Resources
Persons, Events, Architectures... Entities

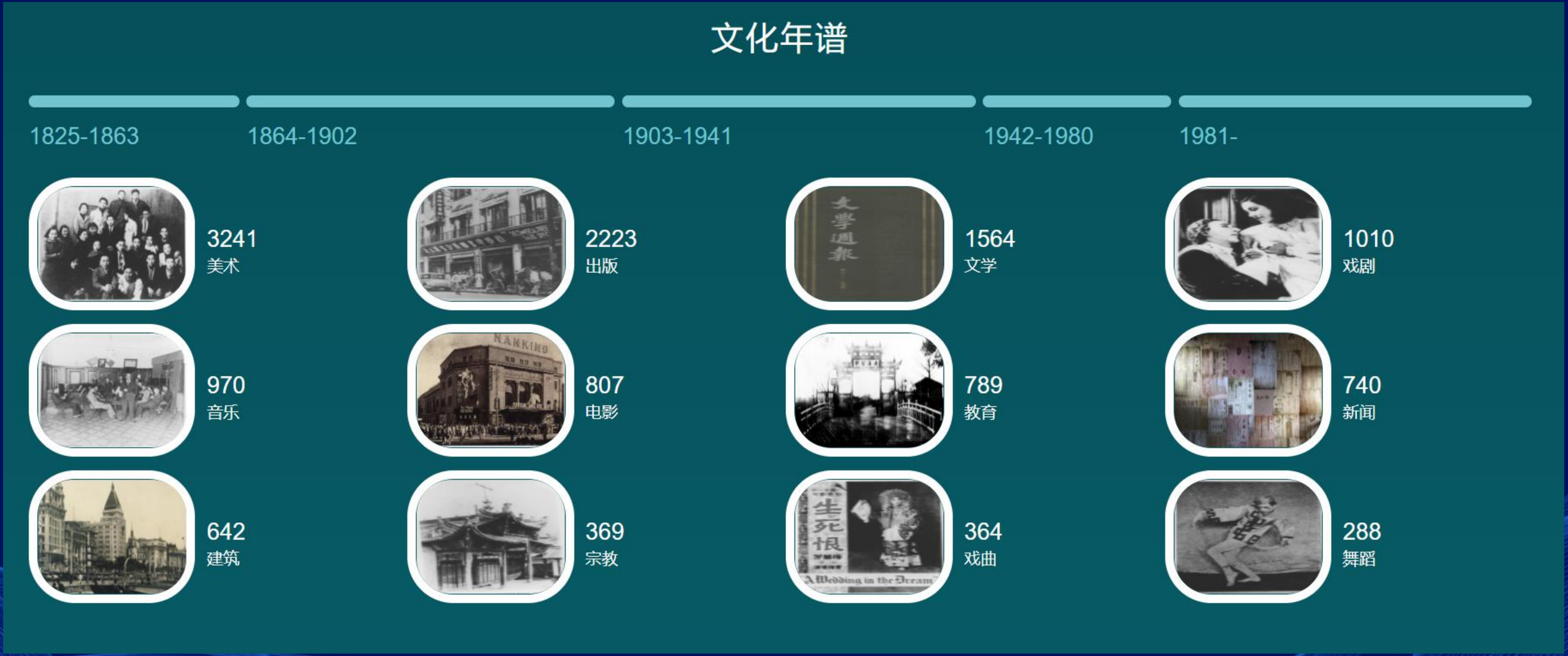
Spatial-temporal Data Infrastructure



http://dhc.library.sh.c

Event Knowledgebase
About the cultural history of Shanghai

http://scc.library.sh.cn (go live soon)



文化年谱

1825-1863 1864-1902 1903-1941 1942-1980 1981-

3241
美术

2223
出版

1564
文学

1010
戏剧

970
音乐

807
电影

789
教育

740
新闻

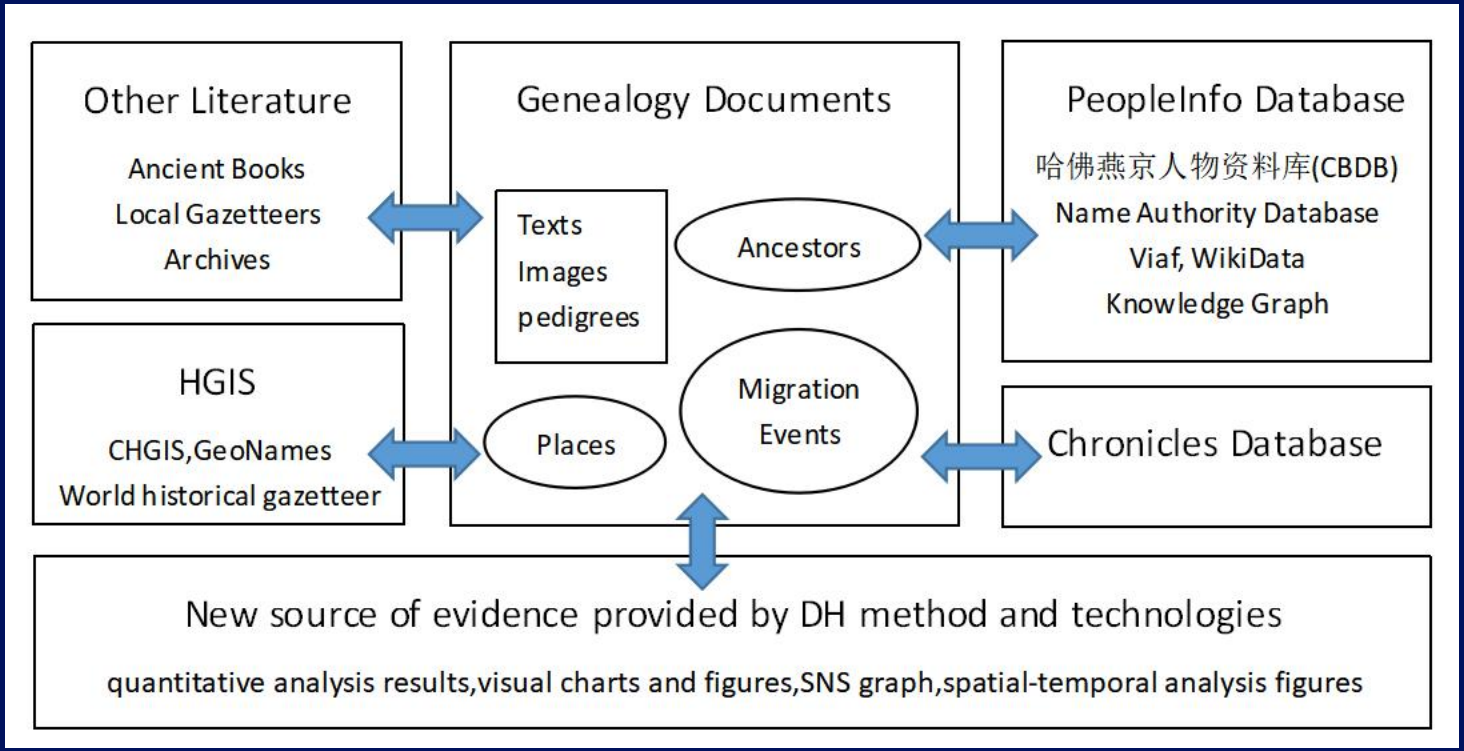
642
建筑

369
宗教

364
戏曲

288
舞蹈

Genealogy Documents as Evidence for Migration History Study



Other Literature

Ancient Books

Local Gazetteers

Archives

HGIS

CHGIS,GeoNames

World historical gazetteer

Genealogy Documents

Texts

Images

pedigrees

Ancestors

Places

Migration Events

PeopleInfo Database

哈佛燕京人物资料库(CBDB)

Name Authority Database

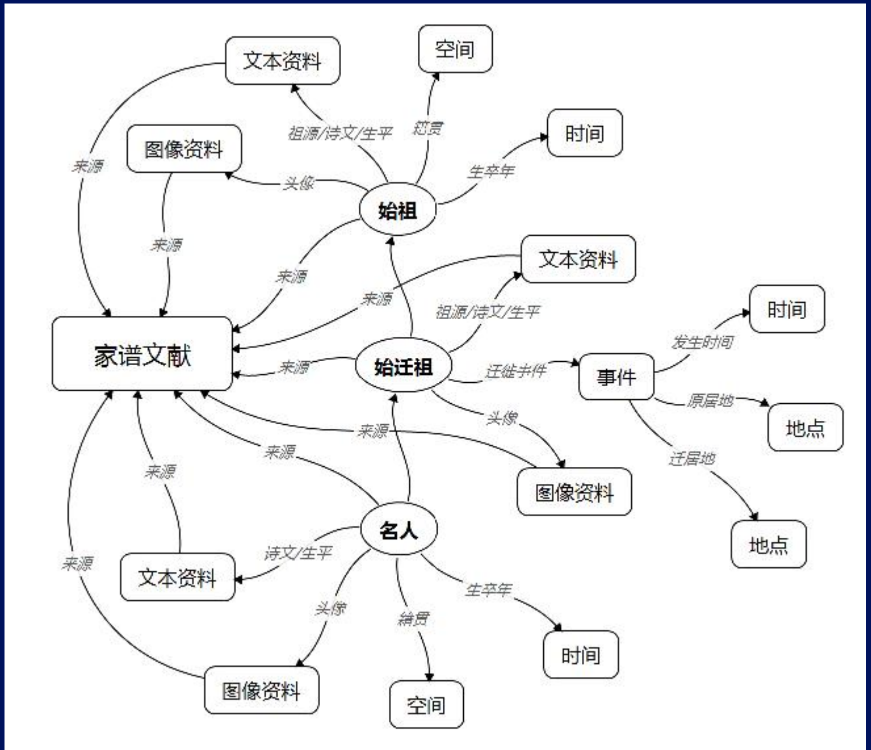
Viaf, WikiData

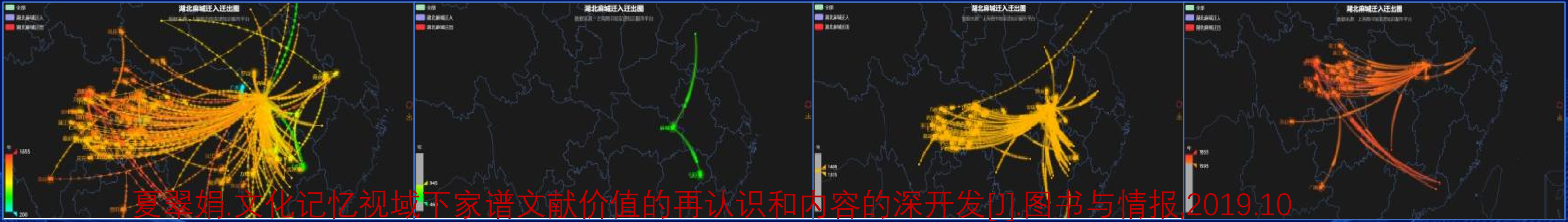
Knowledge Graph

Chronicles Database

New source of evidence provided by DH method and technologies

quantitative analysis results,visual charts and figures,SNS graph,spatial-temporal analysis figures



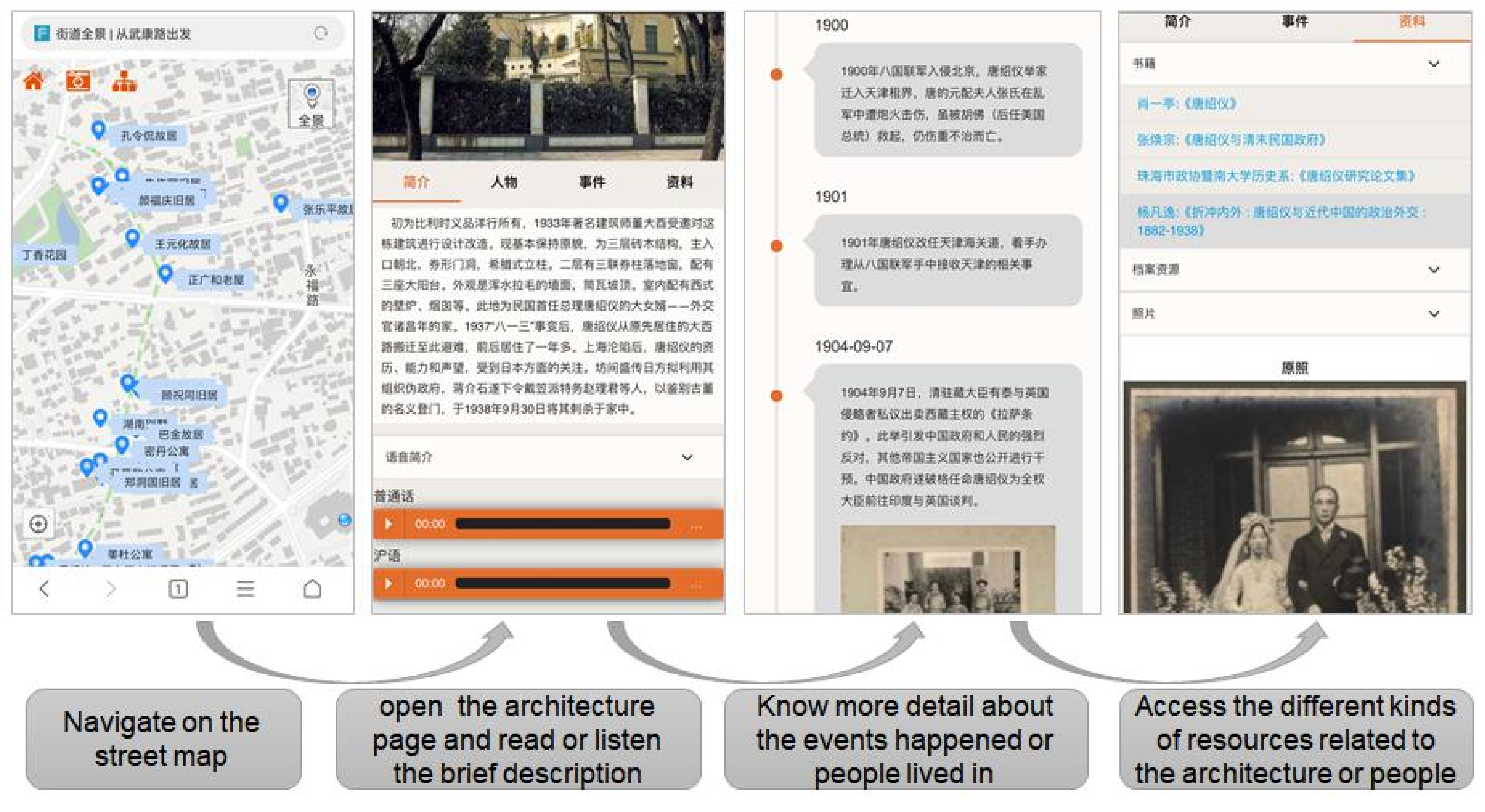


夏翠娟.文化记忆视域下家谱文献价值的再认识和内容的深开发[J].图书与情报,2019.10

data-drven multiple sources of evidence

Cutural Tourisim

http://wkl.library.sh.cn



简介 人物 事件 资料

初为比利时义品洋行所有，1933年著名建筑师董大酉受邀对这栋建筑进行设计改造。现基本保持原貌，为三层砖木结构，主入口朝北，券形门洞，希腊式立柱。二层有三联券柱落地窗，配有三座大阳台。外观是浑水拉毛的墙面，筒瓦坡顶。室内配有西式的壁炉、烟囱等。此地为民国首任总理唐绍仪的大女婿——外交官诸昌年的家。1937"八一三"事变后，唐绍仪从原先居住的大西路搬迁至此避难，前后居住了一年多。上海沦陷后，唐绍仪的资历、能力和声望，受到日本方面的关注，坊间盛传日方拟利用其组织伪政府，蒋介石遂下令戴笠派特务赵理君等人，以鉴别古董的名义登门，于1938年9月30日将其刺杀于家中。

语音简介

普通话

00:00

沪语

00:00

1900

1900年八国联军入侵北京，唐绍仪举家迁入天津租界，唐的元配夫人张氏在乱军中遭炮火击伤，虽被胡佛（后任美国总统）救起，仍伤重不治而亡。

1901

1901年唐绍仪改任天津海关道，着手办理从八国联军手中接收天津的相关事宜。

1904-09-07

1904年9月7日，清驻藏大臣有泰与英国侵略者私议出卖西藏主权的《拉萨条约》。此举引发中国政府和人民的强烈反对，其他帝国主义国家也公开进行干预，中国政府遂破格任命唐绍仪为全权大臣前往印度与英国谈判。

简介 事件 资料

书籍

肖一亭：《唐绍仪》

张焕宗：《唐绍仪与清末民国政府》

珠海市政协暨南大学历史系：《唐绍仪研究论文集》

杨凡逸：《折冲内外：唐绍仪与近代中国的政治外交：1882-1938》

档案资源

照片

原照

Navigate on the street map

open the architecture page and read or listen the brief description

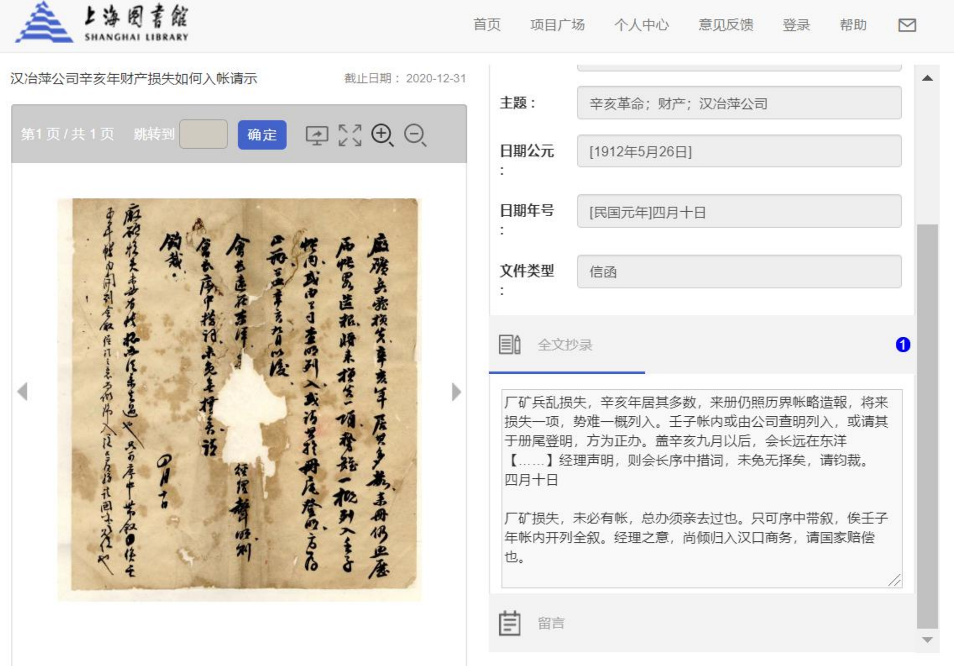
Know more detail about the events happened or people lived in

Access the different kinds of resources related to the architecture or people

Xia Cuijuan,Wang Lihua,Liu Wei. Shanghai memory as a digital humanities platform to rebuild the history of the city. Digital Scholarship in the Humanities, 2021(3).

Crowd-sourcing for UGC (User Generated Content)

- Transcription online
- Captcha



汉冶萍公司辛亥年财产损失如何入帐请示　截止日期：2020-12-31

主题：辛亥革命；财产；汉冶萍公司

日期公元：[1912年5月26日]

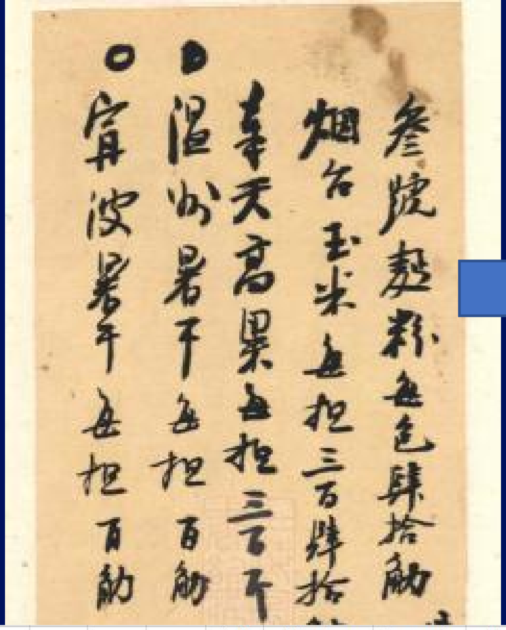
日期年号：[民国元年]四月十日

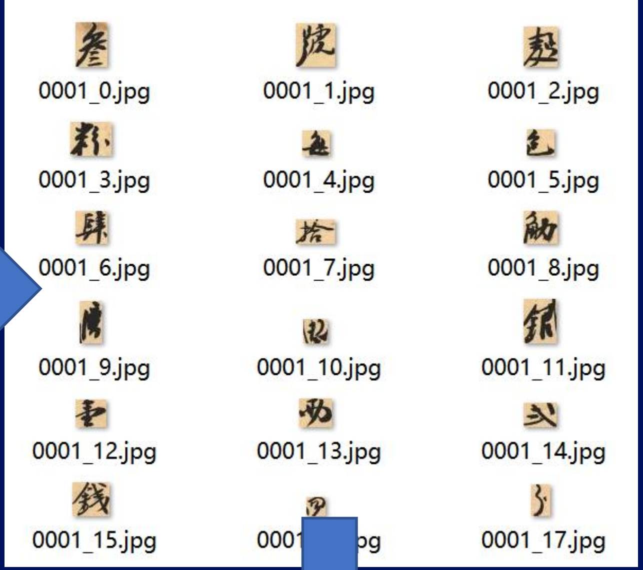
文件类型：信函

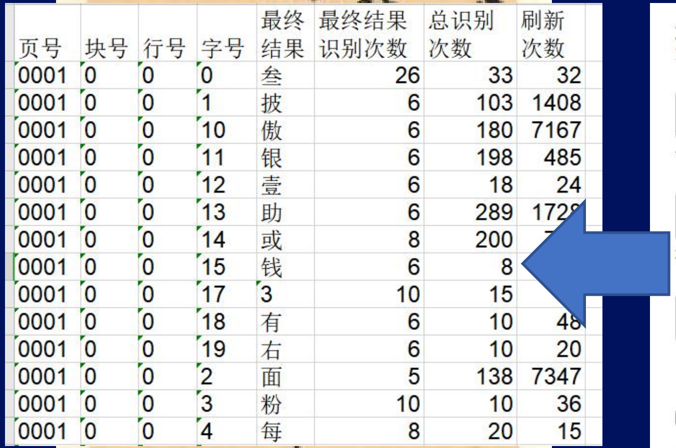
厂矿兵乱损失，辛亥年居其多数，来册仍照历界帐略造報，将来损失一项，势难一概列入。壬子帐内或由公司查明列入，或请其于册尾登明，方为正办。盖辛亥九月以后，会长远在东洋【……】经理声明，则会长序中措词，未免无择矣，请钧裁。
四月十日

厂矿损失，未必有帐，总办须亲去过也。只可序中带叙，俟壬子年帐内开列全叙。经理之意，尚倾归入汉口商务，请国家赔偿也。

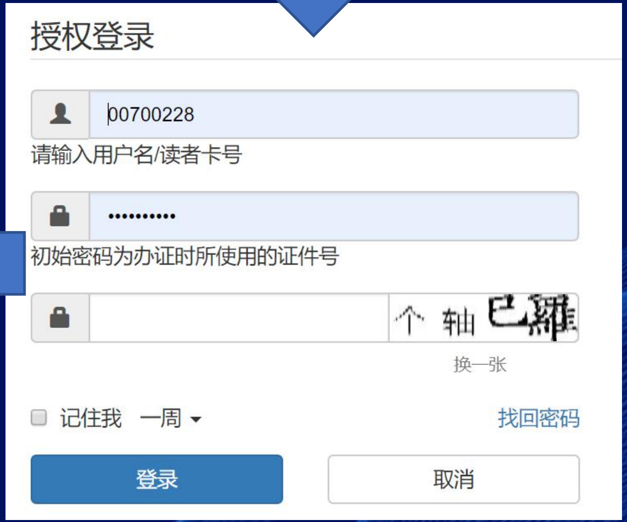
http://zb.library.sh.cn







页号	块号	行号	字号	最终结果	最终结果识别次数	总识别次数	刷新次数
0001	0	0	0	叁	26	33	32
0001	0	0	1	披	6	103	1408
0001	0	0	10	傲	6	180	7167
0001	0	0	11	银	6	198	485
0001	0	0	12	壹	6	18	24
0001	0	0	13	助	6	289	1728
0001	0	0	14	或	8	200	[illegible]
0001	0	0	15	钱	6	8	[illegible]
0001	0	0	17	3	10	15	[illegible]
0001	0	0	18	有	6	10	48
0001	0	0	19	右	6	10	20
0001	0	0	2	面	5	138	7347
0001	0	0	3	粉	10	10	36
0001	0	0	4	每	8	20	15



授权登录

00700228
请输入用户名/读者卡号

••••••••••
初始密码为办证时所使用的证件号

□ 记住我　一周

找回密码

贺晨芝, 张磊. 图书馆数字人文众包项目实践[J]. 图书馆论坛, 2020, 040(005):3-9.

Dublin Core™ Metadata Initiative

Thanks

Cuijuan Xia
xtykc@yeah.net

DCMI 2021
VIRTUAL
Online, 2021 October 4-15