上海图书馆　开放数据应用开发竞赛　2017

名人手稿及档案关联开放数据——内容及结构

上海图书馆高级工程师　夏翠娟

2017. 5. 10

背景与目标

建设人地时事物基本知识库和文献知识库
成为面向人文研究的数据基础设施的一部分

历程

在家谱知识服务平台的基础上，建设盛档、手稿、古籍等文献知识库

家谱原型
2015.1

家谱上线
2016.1

盛档上线
2016.11

手稿档案及古籍原型完成
2017.5

家谱开放数据竞赛
2016.4

手稿档案开放数据竞赛
2017.3

家谱，盛档，手稿，古籍二期
2017.5～

进展

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

家譜 知識服務平台
Genealogy knowledge service platform

http://jp.library.sh.cn

上海图书馆
名人手稿档案库

http://sg.library.sh.cn

盛宣怀档案知識庫

http://sd.library.sh.cn

规范
控制

地

人

时

开放数据竞赛2017

家譜 知識服務平台
Genealogy knowledge service platform

http://jp.library.sh.cn

盛宣怀档案知識庫

http://sd.library.sh.cn

上海图书馆
名人手稿档案库

http://sg.library.sh.cn

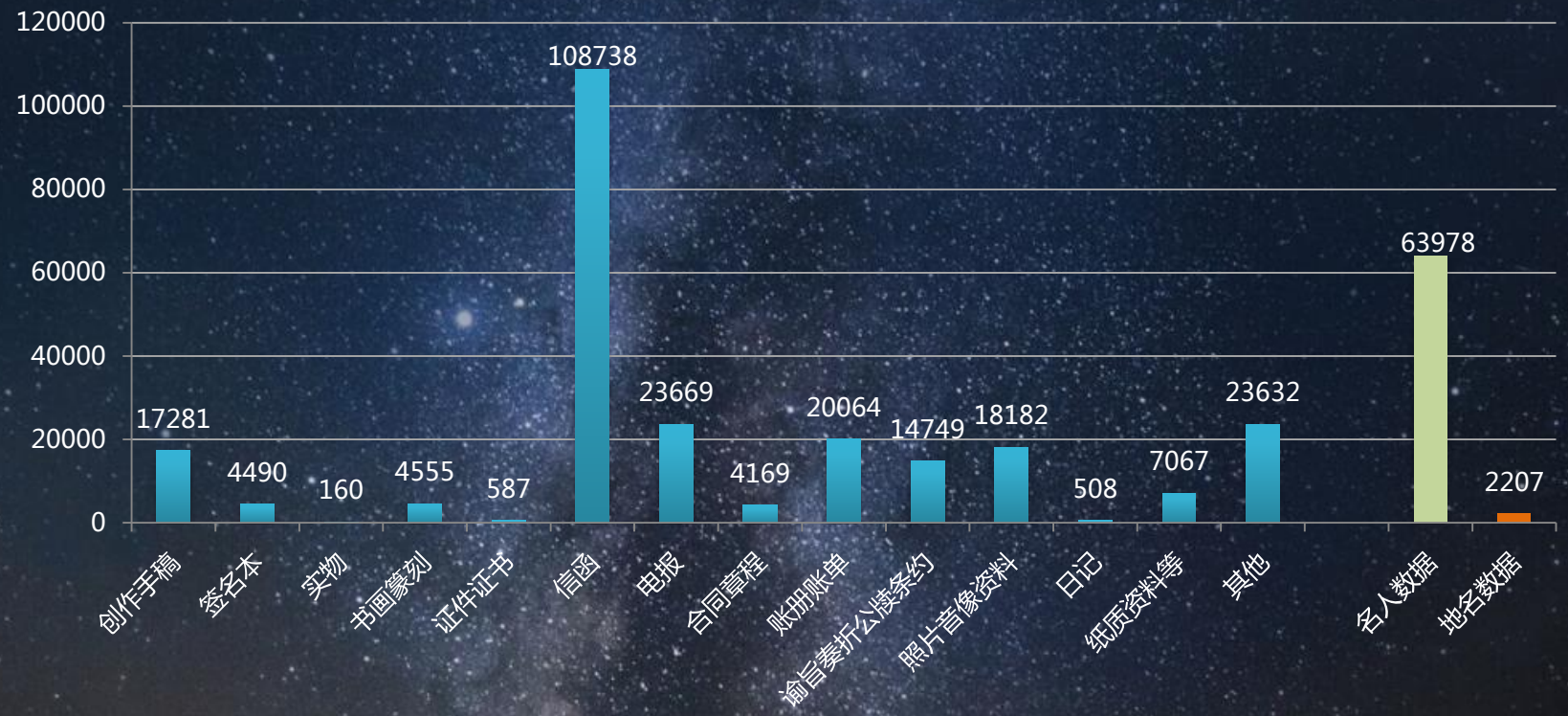
规范
控制

地

人

时

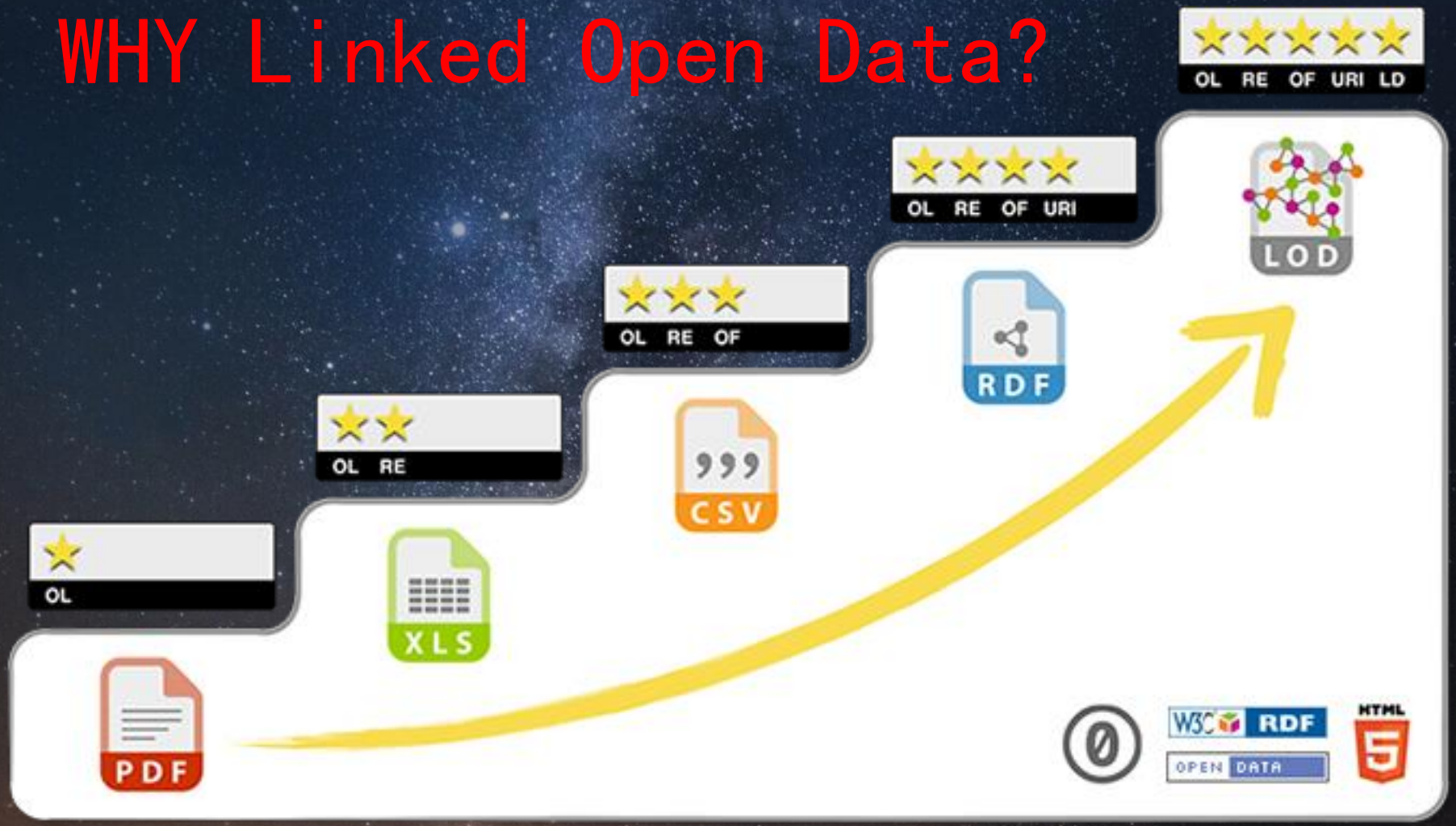
开放数据的内容



共24万余种手稿及档案的元数据（但又不仅是元数据）
涉及名人6万余、地点2千余个（不仅是人名和地名）
时间跨度为19世纪初至20世纪末近200年

开放数据的技术——LOD

WHY Linked Open Data?



什么是关联数据

用HTTP URI作为一切事物的名称

当访问HTTP URI时提供RDF数据

尽可能多地描述事物间的关系并使机器可理解

原则一、三

用HTTP URI作为一切事物的名称

手稿：

http://data.library.sh.cn/sg/resource/work/h3sv7z31utc53yvp

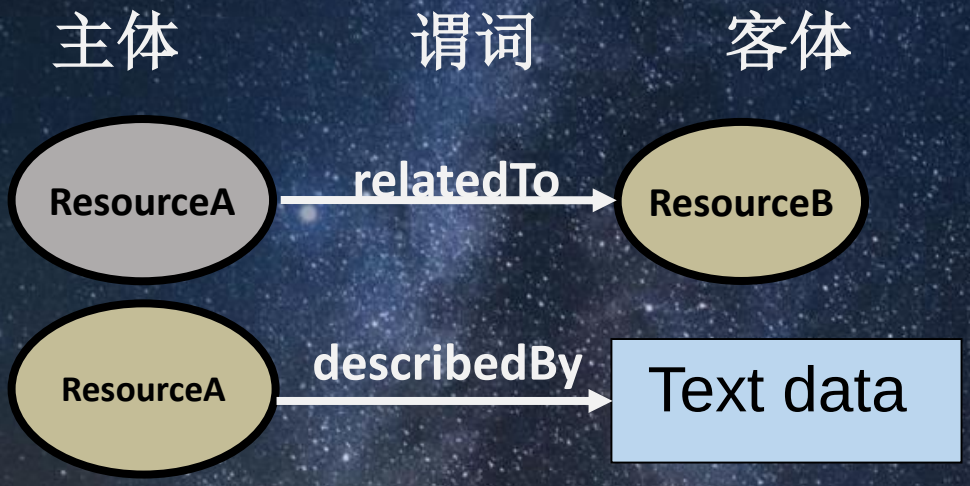
人： http://data.library.sh.cn/entity/person/t3qypozz7y13mfdt

地： http://data.library.sh.cn/entity/place/tk5s4pej6linq9tr

时： ttp://data.library.sh.cn/authority/temporal/4alljneqiivh5691

原则三——

当访问HTTP URI时提供RDF数据

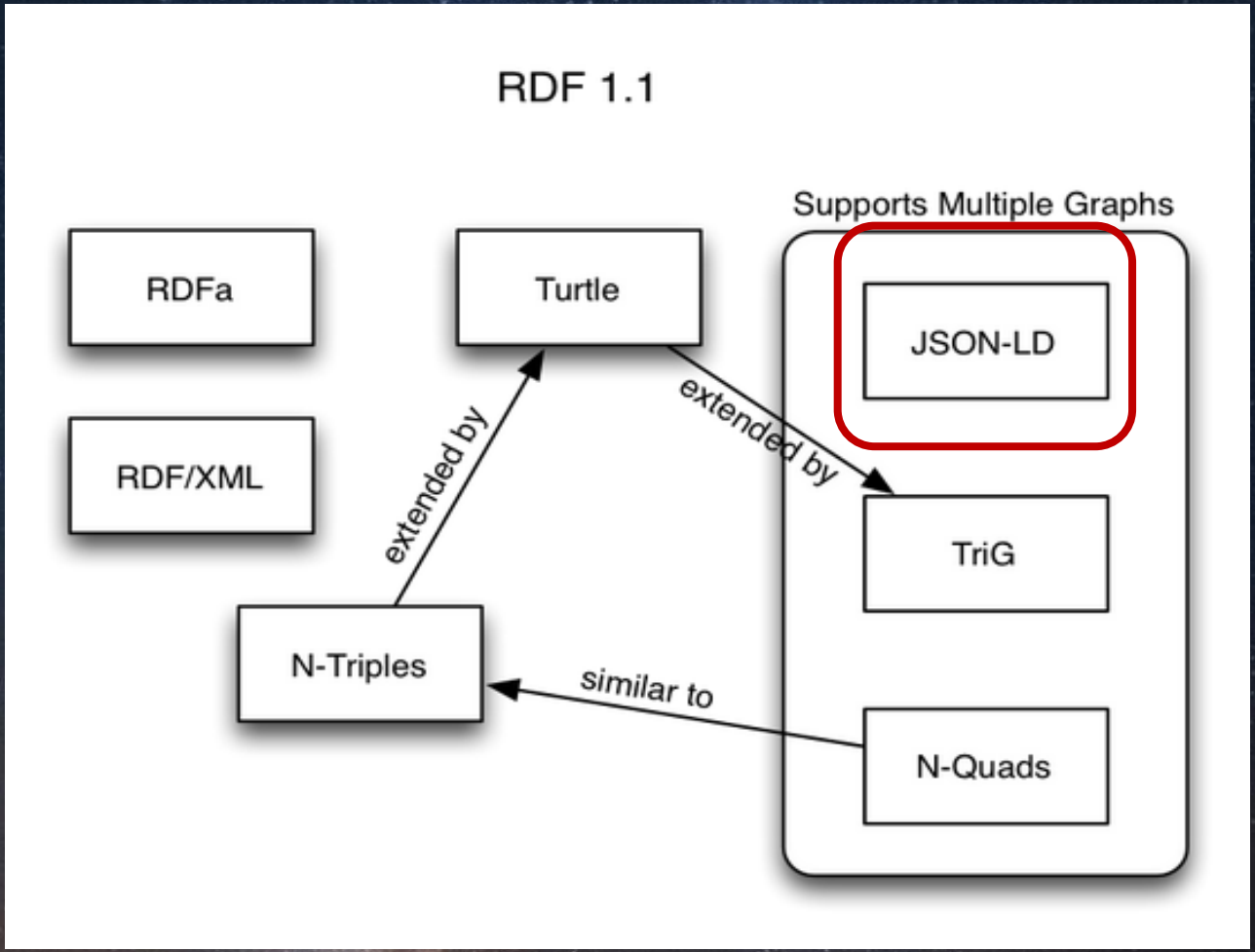


抽象模型
Abstract Model

```
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
<http://data.library.sh.cn/jp/entity/person/etrd44w3m3g1vncn> a
shl:Person;
        foaf:familyName
<http://data.library.sh.cn/authority/familyname/68n959cf8zdfkz3v>;
        foaf:name “夏翠娟”@zh_cn.
```

序列化格式
Serialization

RDF序列化格式



http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/

茅盾

作家 | 1896 - 1981

国籍：中国 | 籍贯：桐乡

民族：汉族 | 性别：男

职衔

异名

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

关系

曹辛之(... 黎澍(友) 赵清阁(...
姚雪垠(... 夏衍(友) 刘白羽(...
萨空了(... 乐黛云(... 端木蕻良..
熊佛西(... 楼适夷(... 吴祖光(...
聂华苓(... 冯乃超(... 冯亦代(...

著名作家,文化活动家.原名沈德鸿,字雁冰.浙江桐乡人.1916年毕业于北京大学预科,入上海商务印书馆编译所工作.曾参加"五四"运动和早期共产主义运动.1921年与郑振铎,叶圣陶等创办文学研究会,主编《小说月报》,同年参加中国共产党.1926年春去广州,任国民党中宣部秘书.后又到武汉,任《民国日报》总编辑."四·一二"政变,宁汉合流之后,被迫回到上海,开始写《幻灭》,《动摇》,《追求》三部曲.1928年夏离沪赴日,同党失去组织关系.1930年回沪,参加左联,写了《子夜》等小说.抗战爆发之后积

```
{
  "@id": "http://data.library.sh.cn/entity/person/t3qypozz7y13mfdt",
  "@type": "http://www.library.sh.cn/ontology/Person",
  "identifier": "A000005",
  "nationality": "http://data.library.sh.cn/entity/nationality/china",
  "nativePlace": "http://data.library.sh.cn/entity/place/l1khsjyerlopeqmp",
  "birthday": "1896",
  "deathday": "1981",
   "speciality": "作家",
  "gender": "男",
  "http://xmlns.com/foaf/0.1/name": [
     { "@language": "cht", "@value": "茅盾" },
     { "@language": "chs", "@value": "茅盾" },
     { "@language": "en", "@value": "maodun" }
  ],
  "ethnicity": "http://data.library.sh.cn/entity/ethnicity/han",
 "name": [
     "http://data.library.sh.cn/entity/nameOther/6clvn1d0hgxvr42c",
     "http://data.library.sh.cn/entity/nameOther/rj7k7ok6tchfw7y8",
     "friendOf": [
     "http://data.library.sh.cn/entity/person/ekw1al118gjv96qy",
     "http://data.library.sh.cn/entity/person/274zyh3p59lt76ye",
      … …],
}
```

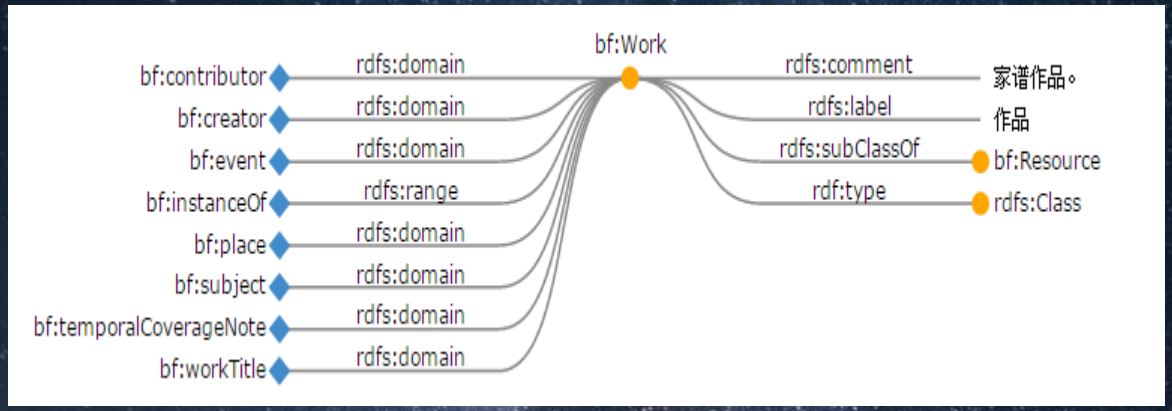
原则四

尽可能地描述事物间关系
并使机器可理解

本体（Ontology）

作品 机构 人 地

领域的可共享的概念模型



术语词表
类、属性（关系）

RDF Schema 1.1 SKOS OWL

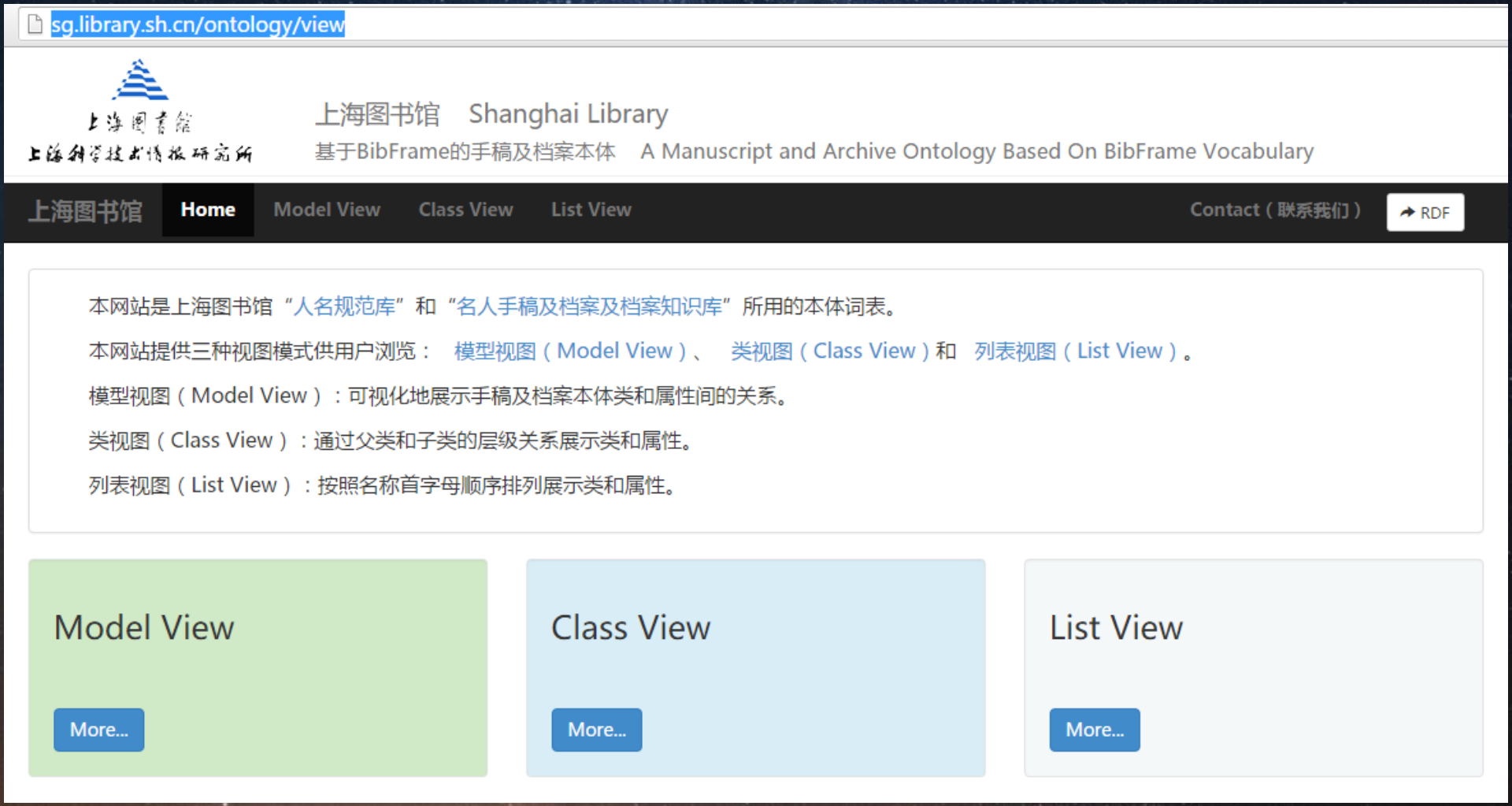
形式化的

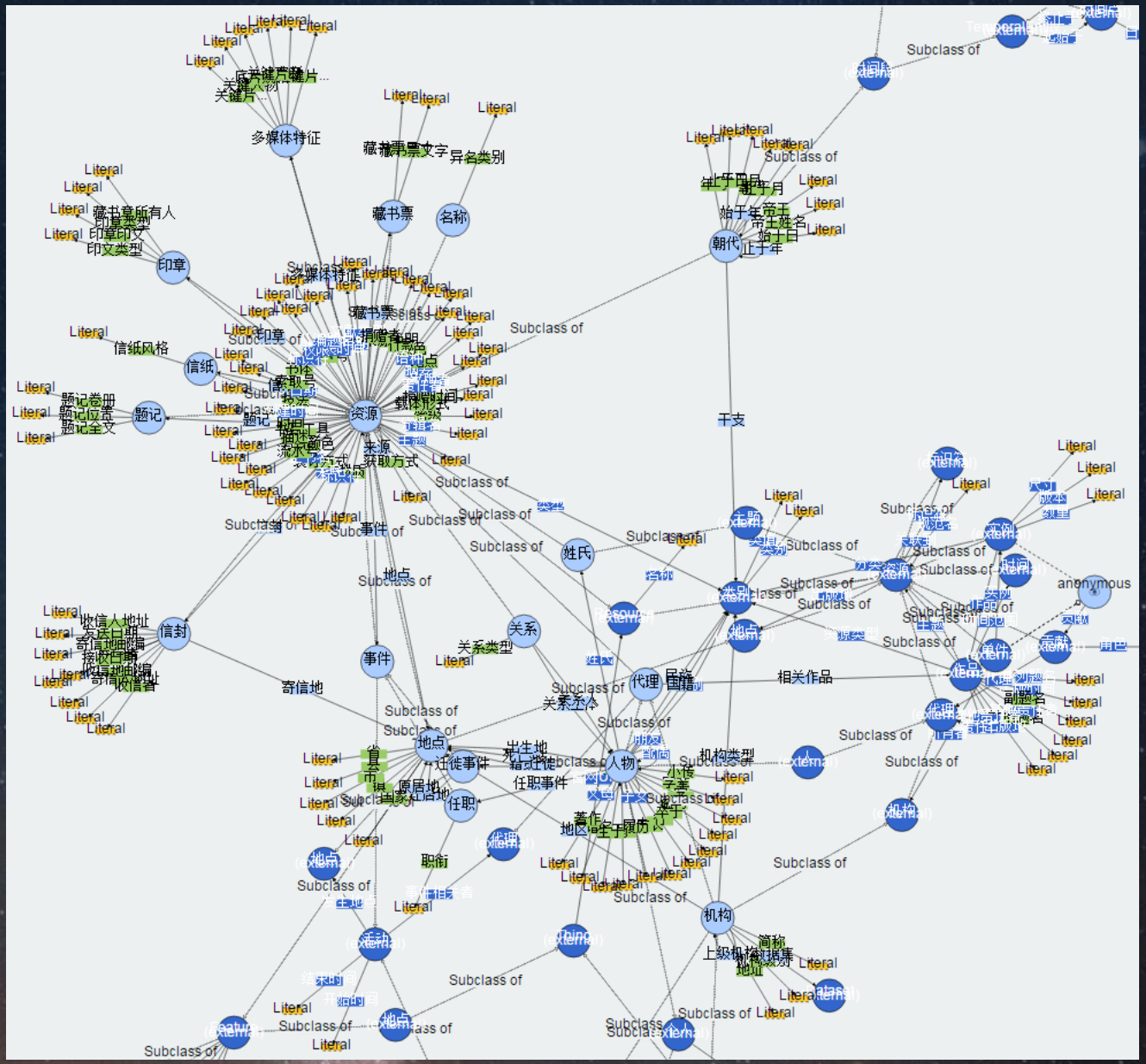
```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix bf: <http://bibframe.org/vocab/> .
bf:Work a rdfs:Class ;
        rdfs:label "作品" ;
        rdfs:subClassOf bf:Resource ;
        rdfs:comment "家谱作品。" .
```

本体(Ontology)是对领域知识进行抽象的、形式化的概念模型。

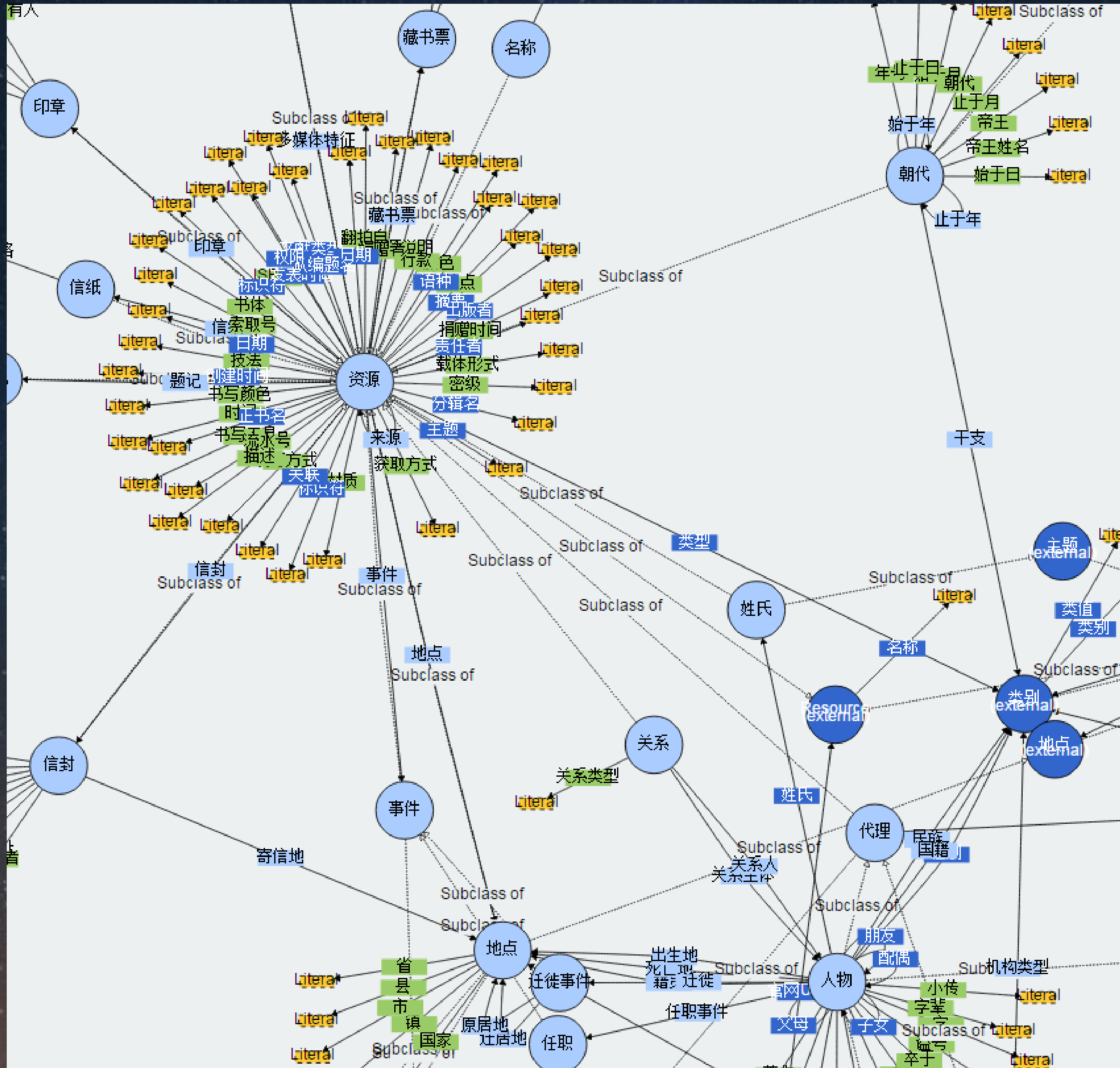
人名规范及手稿档案本体

http://sg.library.sh.cn/ontology/

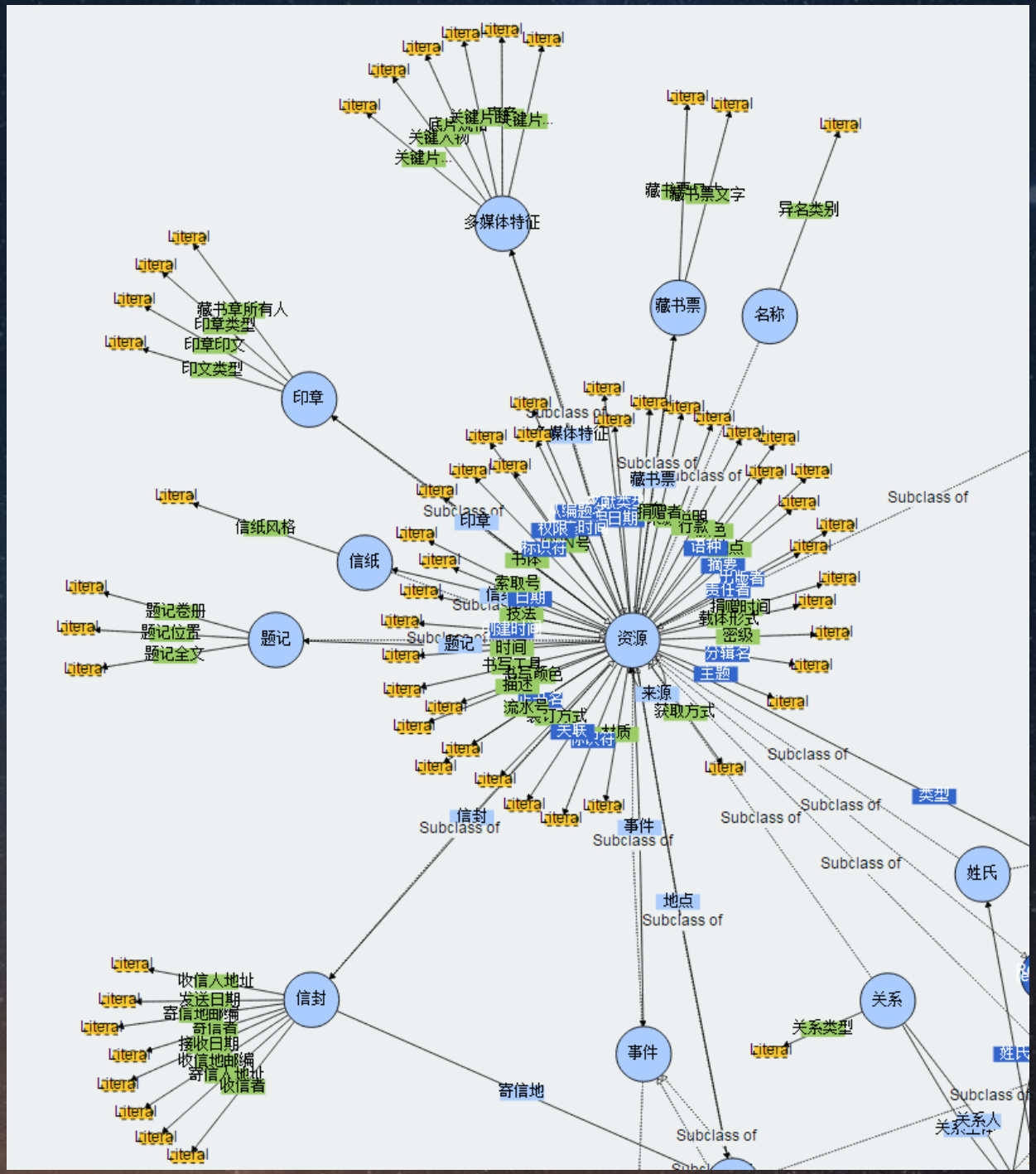


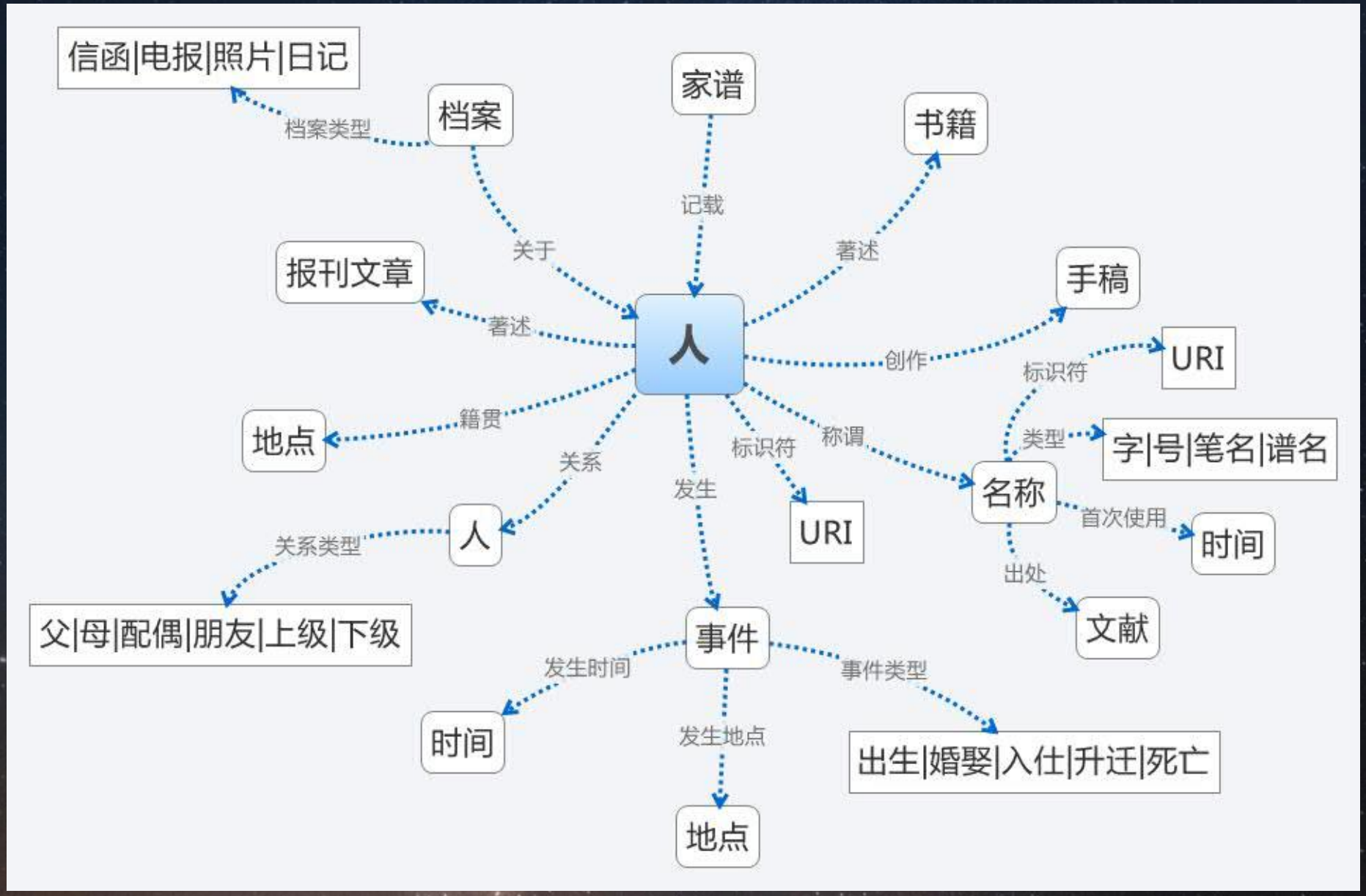


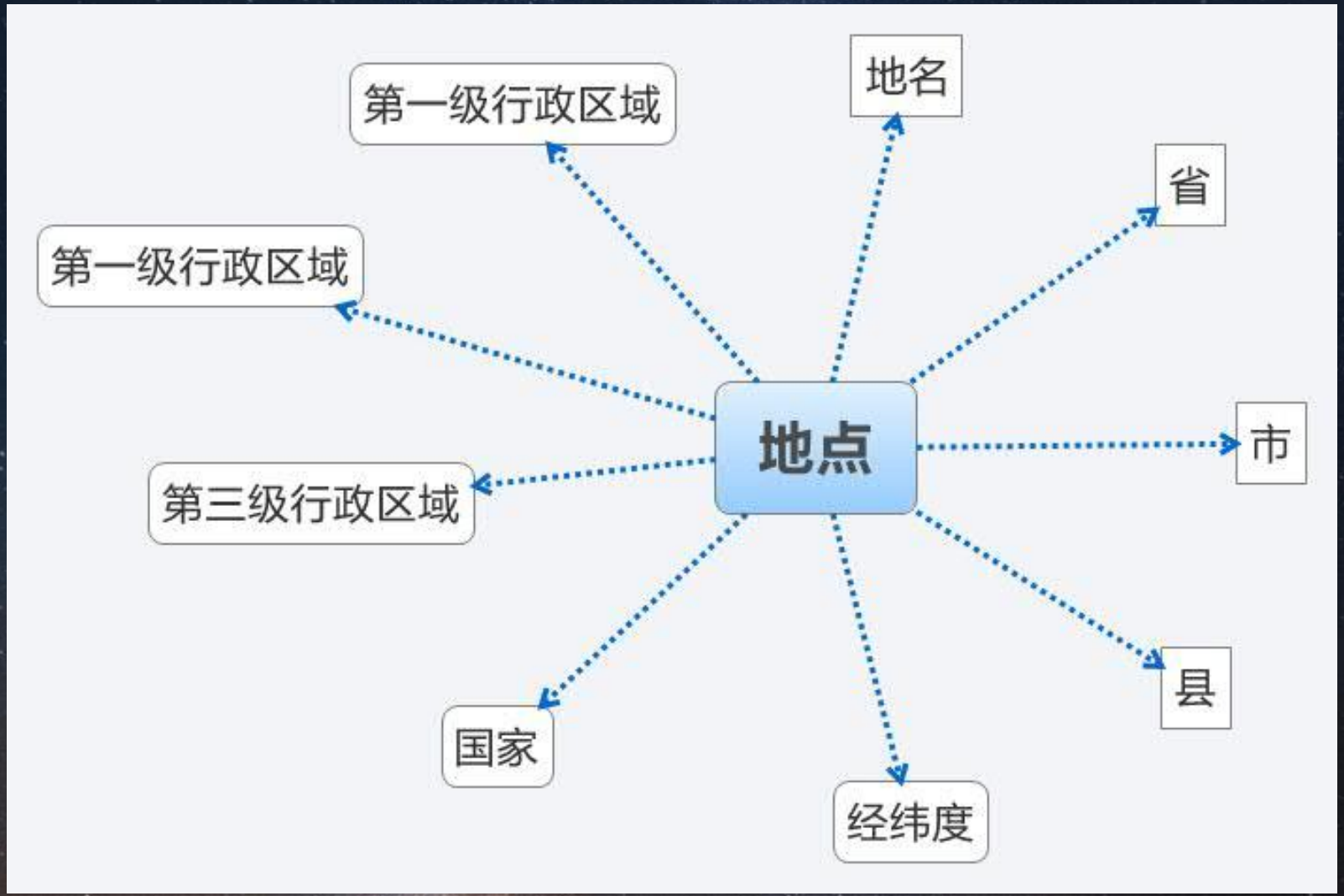
http://usources.cn/full.html#iri=http://www.usources.cn/concept/library_sh_cn

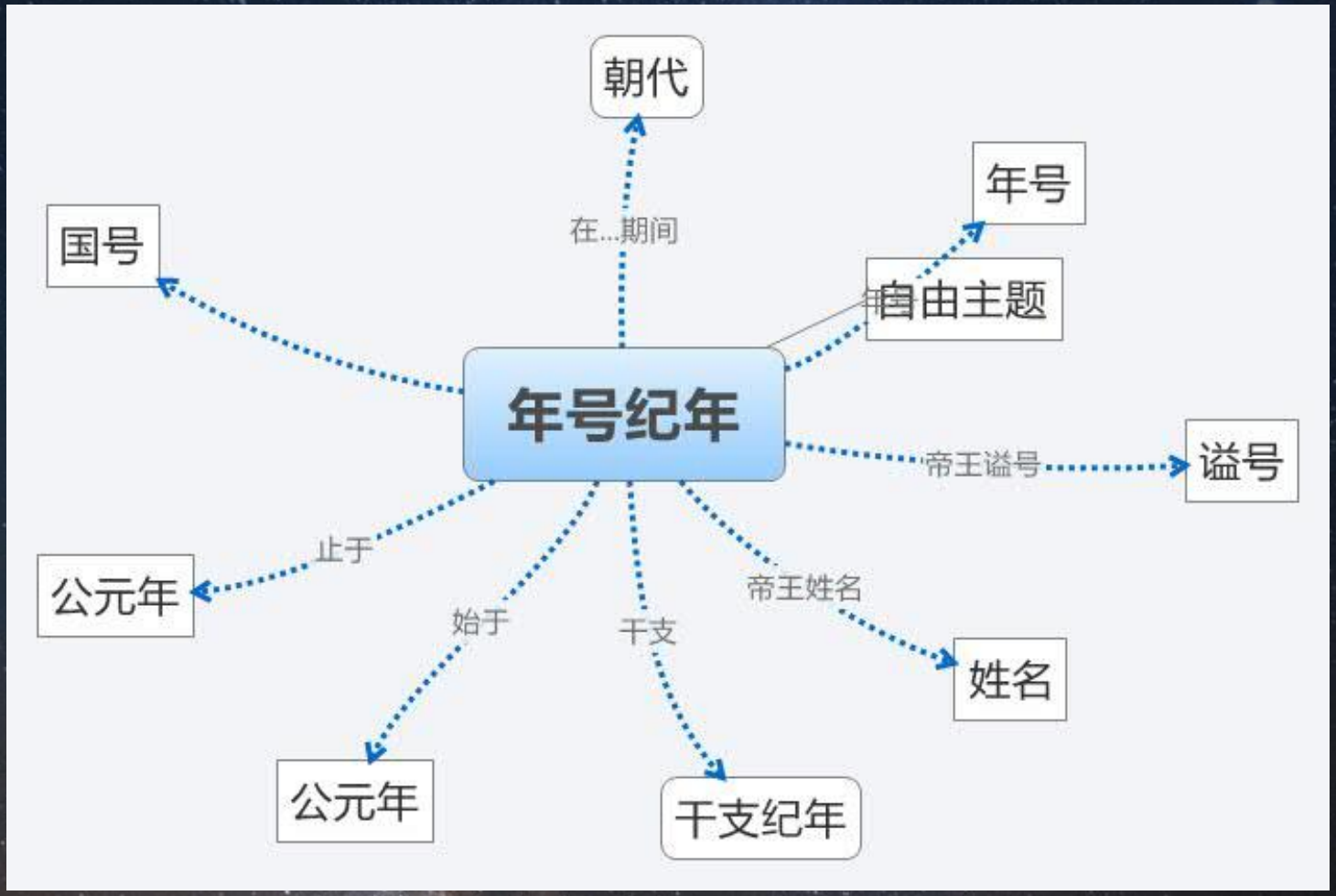


资源
印章
藏书票
名称
朝代
信纸
信封
事件
关系
姓氏
代理
地点
任职
迁徙事件
人物
主题 external
类别 external
地点 external
Resource external
Subclass of
Literal
多媒体特征
书体
日期
技法
题记
书写颜色
来源
获取方式
描述
语种
摘要
责任者
载体形式
密级
分辑名
主题
类型
名称
干支
始于年
止于年
止于月
始于日
帝王
帝王姓名
关系类型
寄信地
出生地
任职事件
原居地
省
县
市
镇
国家
朋友
配偶
小传
字辈
父母
子女
著作
民族
国籍
类值
类别
机构类型


资源
多媒体特征
藏书票
名称
异名类别
印章
藏书章所有人
印章类型
印章印文
印文类型
信纸
信纸风格
题记
题记卷册
题记位置
题记全文
信封
收信人地址
发送日期
寄信地邮编
寄信者
接收日期
收信地邮编
寄信人地址
收信者
寄信地
事件
地点
关系
关系类型
关系人
姓氏
类型
Subclass of
Literal


信函|电报|照片|日记
档案类型
档案
家谱
书籍
记载
关于
著述
报刊文章
著述
人
手稿
创作
URI
标识符
籍贯
地点
关系
发生
标识符
称谓
类型
字|号|笔名|谱名
名称
首次使用
时间
URI
人
关系类型
出处
文献
父|母|配偶|朋友|上级|下级
事件
发生时间
事件类型
时间
发生地点
出生|婚娶|入仕|升迁|死亡
地点


地点
第一级行政区域
地名
省
第一级行政区域
市
第三级行政区域
县
国家
经纬度


朝代
在...期间
年号
国号
自由主题
年号纪年
帝王谥号
谥号
止于
公元年
帝王姓名
始于
干支
姓名
公元年
干支纪年

开放数据竞赛2017—工具

数据开放方式

家譜 知識服務平台
Genealogy knowledge service platform

http://jp.library.sh.cn

盛宣怀档案知識庫

http://sd.library.sh.cn

上海图书馆
名人手稿档案库

http://sg.library.sh.cn

机构名录　地理名词表

中国历史纪年表

http://data.library.sh.cn

SPARQL Endpoint

Restful API

HTTP URI

数据消费接口

JSO
N-LD

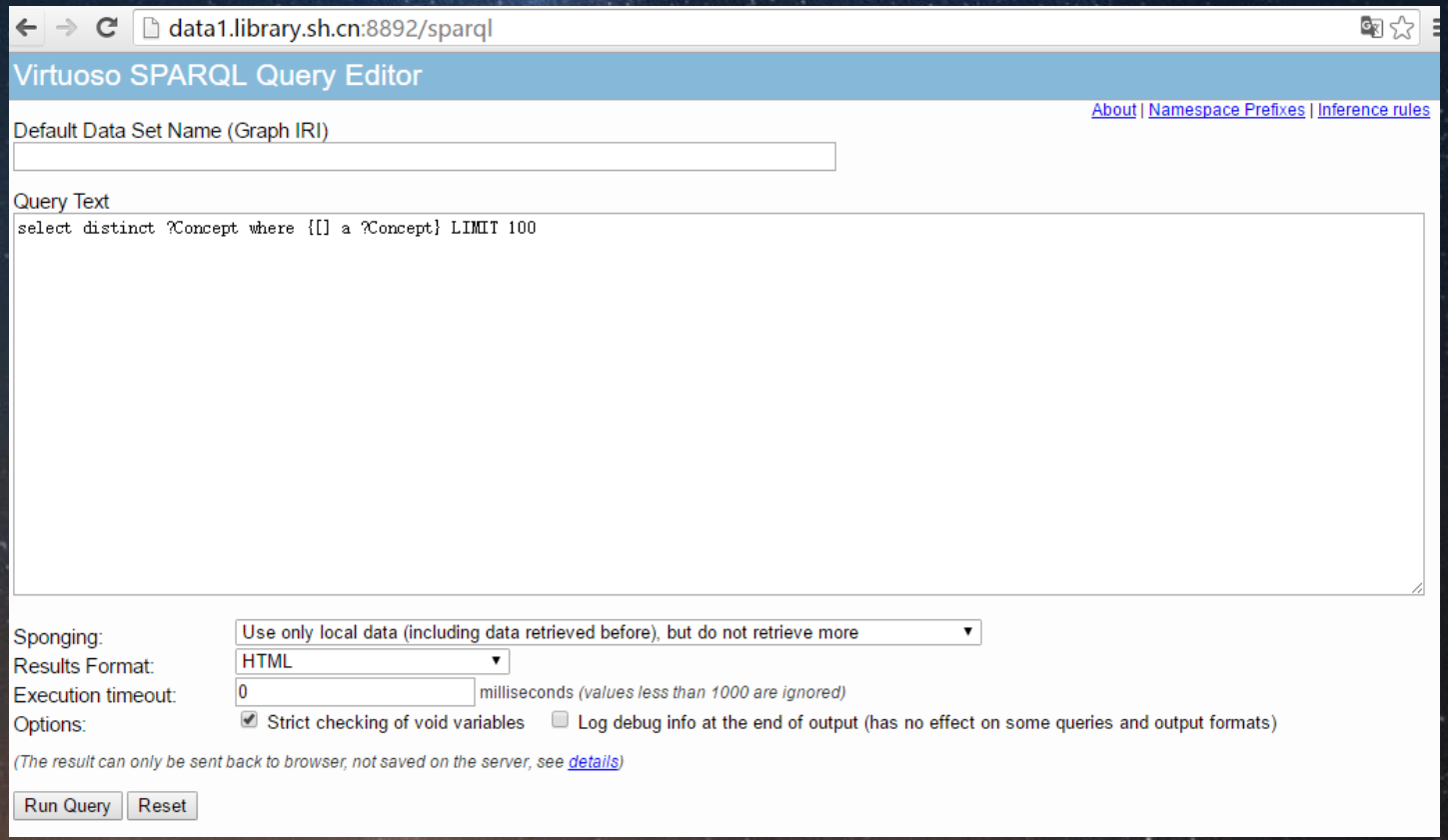
返回数据

Sparql Endpoint

家谱：http://data1.library.sh.cn:8890/sparql

盛档：http://data1.library.sh.cn:8892/sparql

名人手稿：http://data1.library.sh.cn:8893/sparql

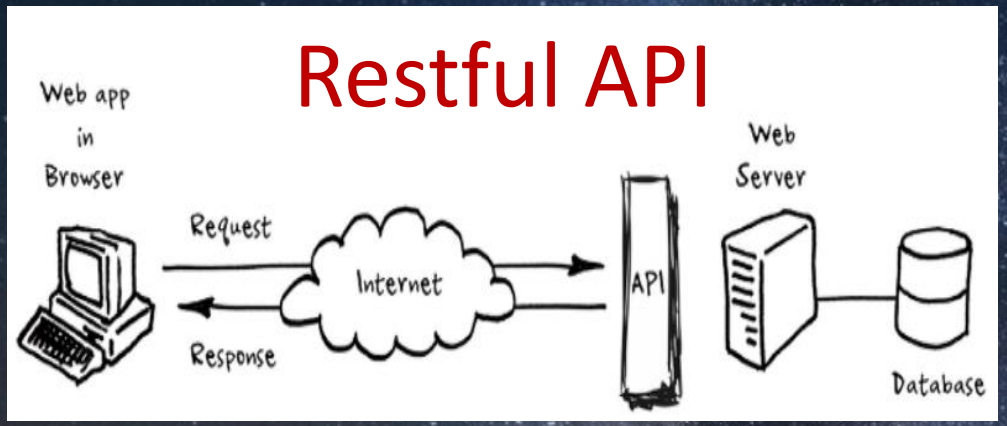


Jav
a

C++

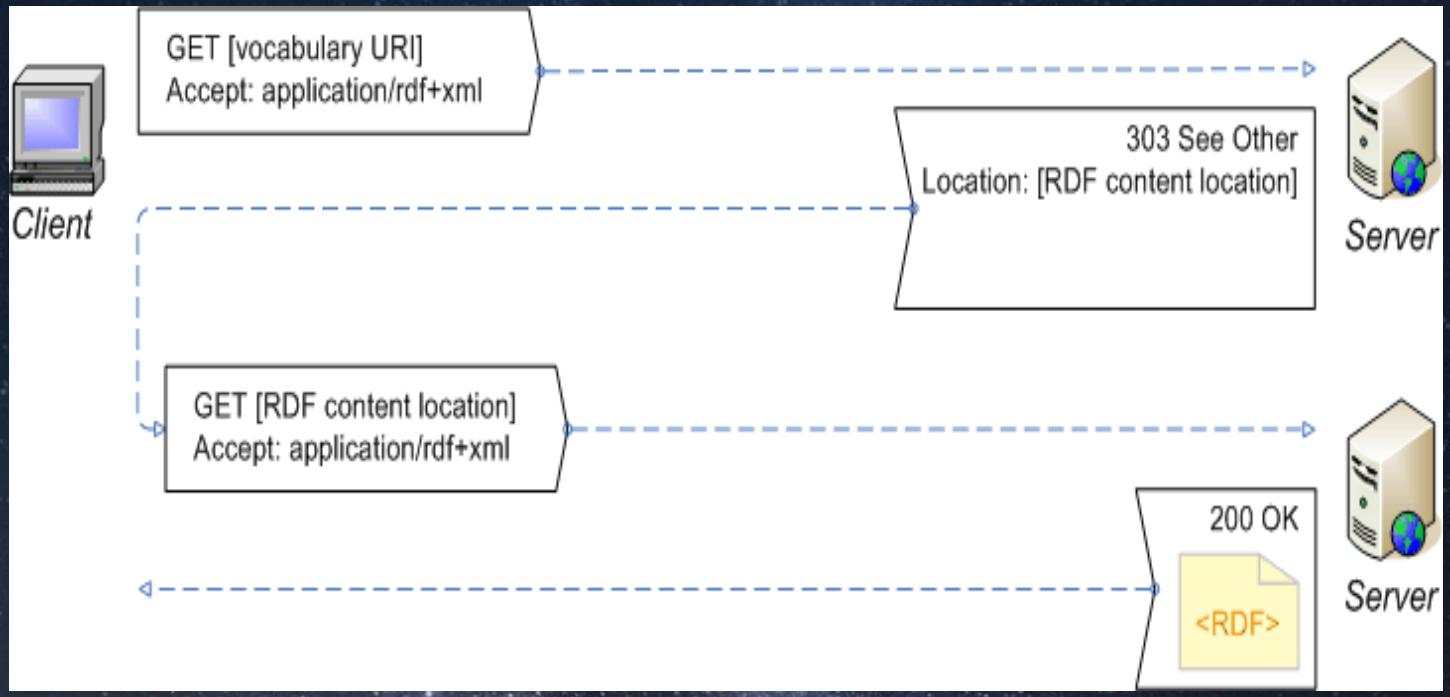
PH
P

Pyth
on



http://data1.library.sh.cn/sg/data/json?uri=[参数1]?key=[参数2]key=YourAPIKey

经过封装进行了数据整合



当客户端访问URI时，服务器端需要根据其发送请求的HTTP报头返回合适的表示形式（如HTML、RDF、PDF 等）。

一起来，更精彩！

Be together, Be better!

祝各位取得好成绩！

cjxia@libnet.sh.cn