

上海图书馆 开放数据应用开发竞赛 2017

上海图书馆开放数据应用 开发竞赛技术方案分享

上海图书馆高级工程师 夏翠娟

2017.10.27

背景



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>

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-家谱



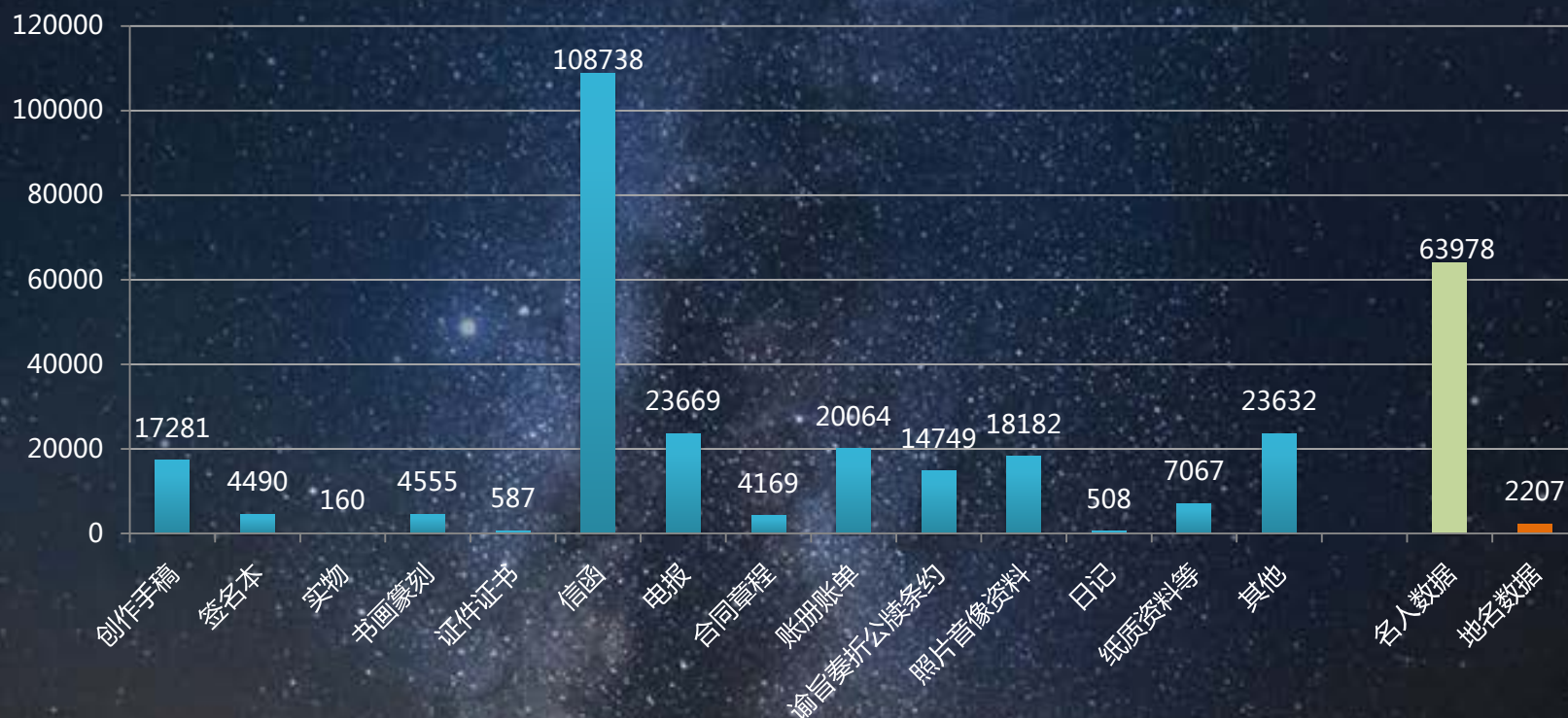
<http://jp.library.sh.cn>

开放数据的内容-世系表

先祖名人
共1.4万余
人



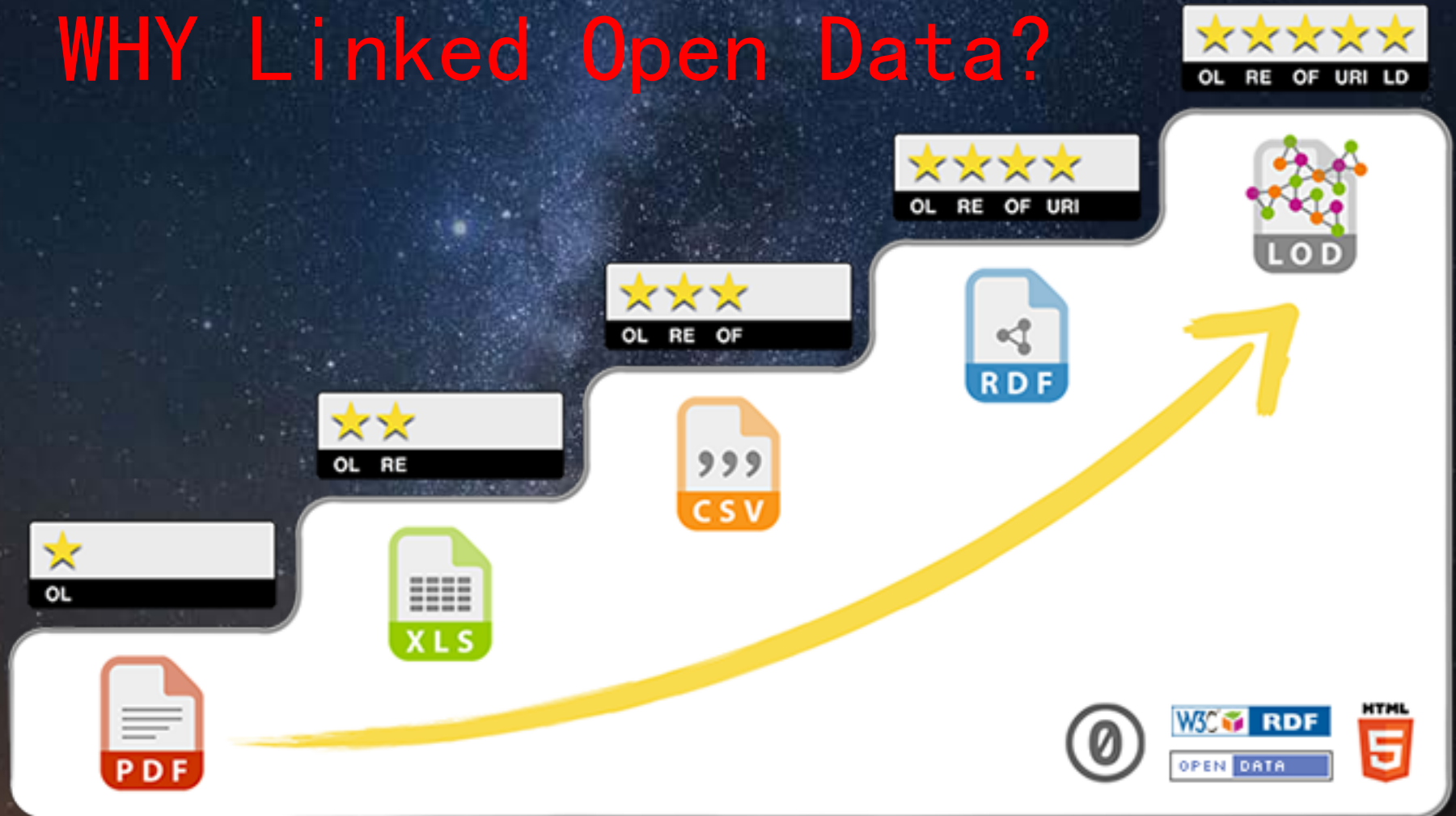
开放数据的内容-手稿档案



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

< URI

HTTP
URI

RDF

关系 >

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、二

用HTTP URI作为一切事物的名称

手稿:

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

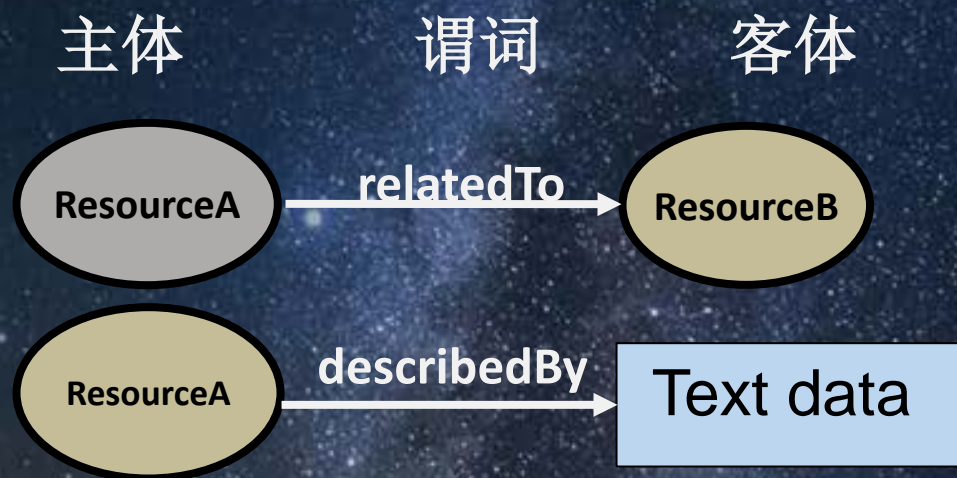
人: <http://data.library.sh.cn/entity/person/t3qypo7y13mfdt>

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

时: <http://data.library.sh.cn/authority/temporal/4alljneqiivh5691>

原则三——

当访问HTTP URI时提供RDF数据

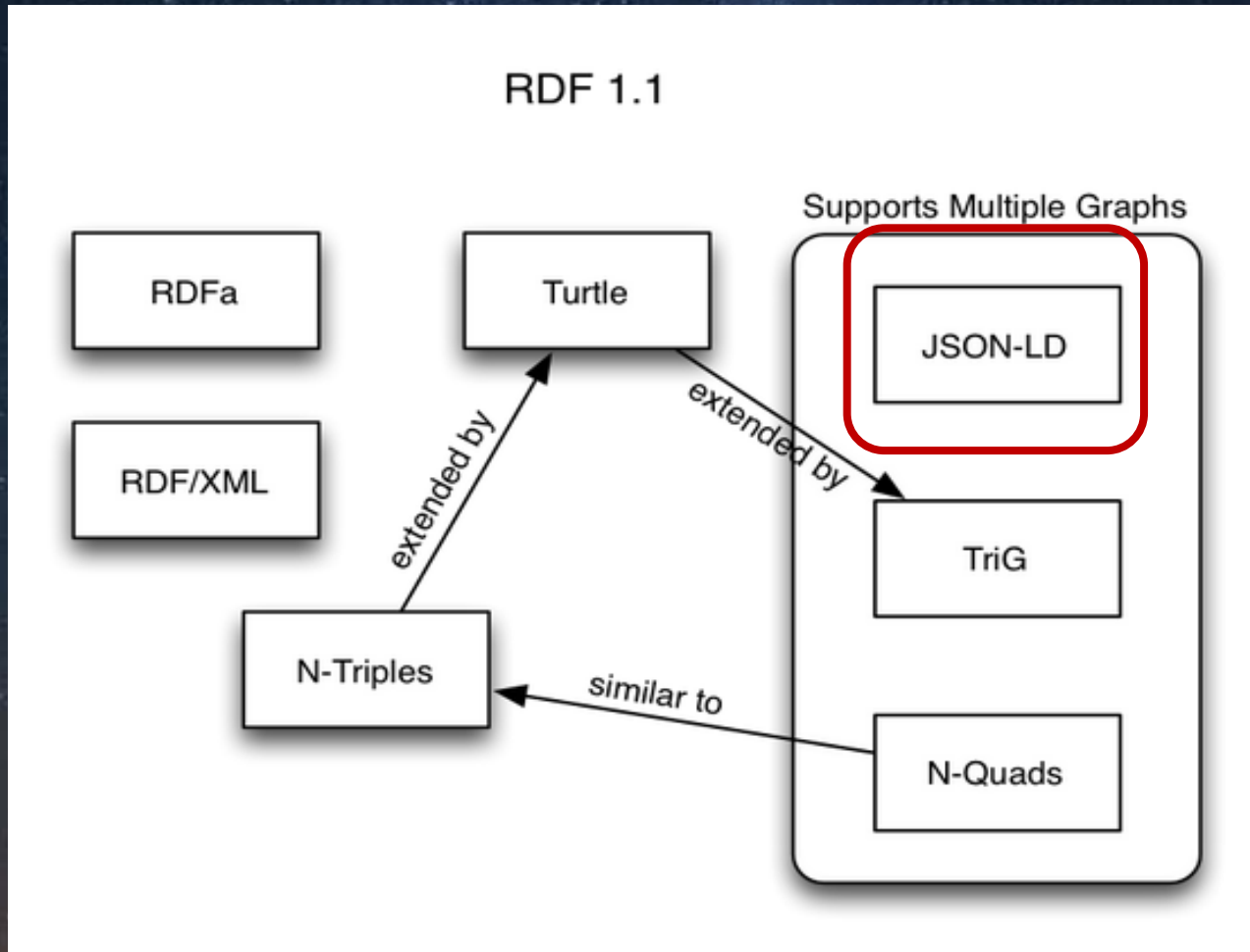


抽象模型
Abstract
Model

```
@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



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

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(…


```
{
  "@id": "http://data.library.sh.cn/entity/person/t3qypozz7y13mfddt",
  "@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/ekw1al118gjb96qy",
      "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）

人名规范及手稿档案本体

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

sg.library.sh.cn/ontology/view



上海图书馆

上海科学技术情报研究所

上海图书馆 Shanghai Library

基于BibFrame的手稿及档案本体 A Manuscript and Archive Ontology Based On BibFrame Vocabulary

上海图书馆

Home

Model View

Class View

List View

Contact (联系我们)

[RDF](#)

本网站是上海图书馆“人名规范库”和“名人手稿及档案及档案知识库”所用的本体词表。

本网站提供三种视图模式供用户浏览：[模型视图 \(Model View\)](#)、[类视图 \(Class View\)](#)和 [列表视图 \(List View\)](#)。

模型视图 (Model View)：可视化地展示手稿及档案本体类和属性间的关系。

类视图 (Class View)：通过父类和子类的层级关系展示类和属性。

列表视图 (List View)：按照名称首字母顺序排列展示类和属性。

Model View

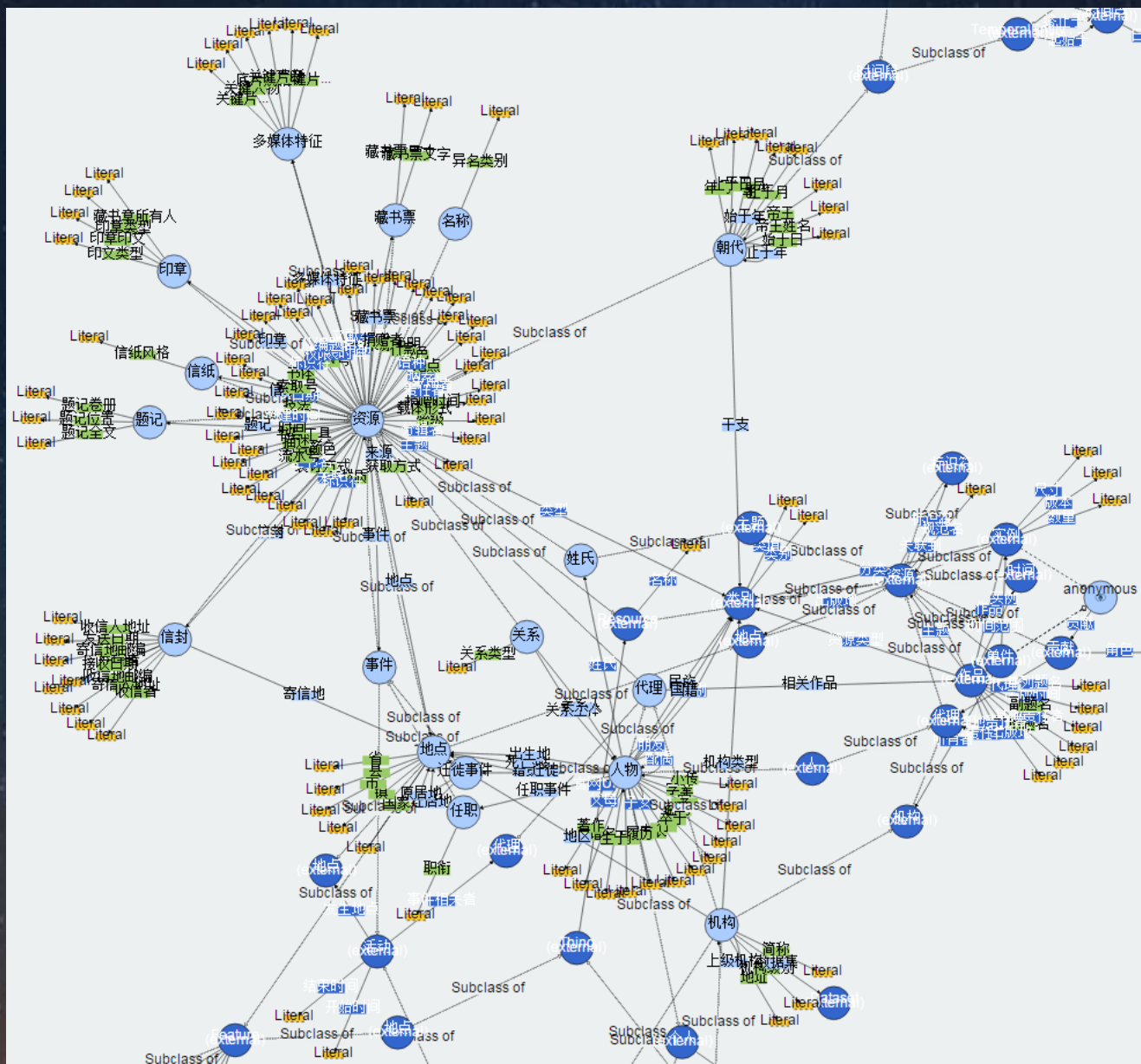
[More...](#)

Class View

[More...](#)

List View

[More...](#)



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

开放数据竞赛2017—工具

data1.library.sh.cn/tools/



本网站是上海图书馆的开放数据平台，将陆续以多种数据清洗和转换工具等。提供各种数据消费接口供

RDB2RDF

More...

data1.library.sh.cn/tools/rdb2rdf/

wy8wyG
R2R

从关系数据库到关联数据
国家自然科学基金青年项目13CTQ008成果之一

上传配置

数据库初始化

数据库类型

Mysql

数据库地址

127.0.0.1

数据库端口

3306

数据库名称

Genealogy

数据库用户

root

数据库密码

连接数据库

数据开放内容及方式

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

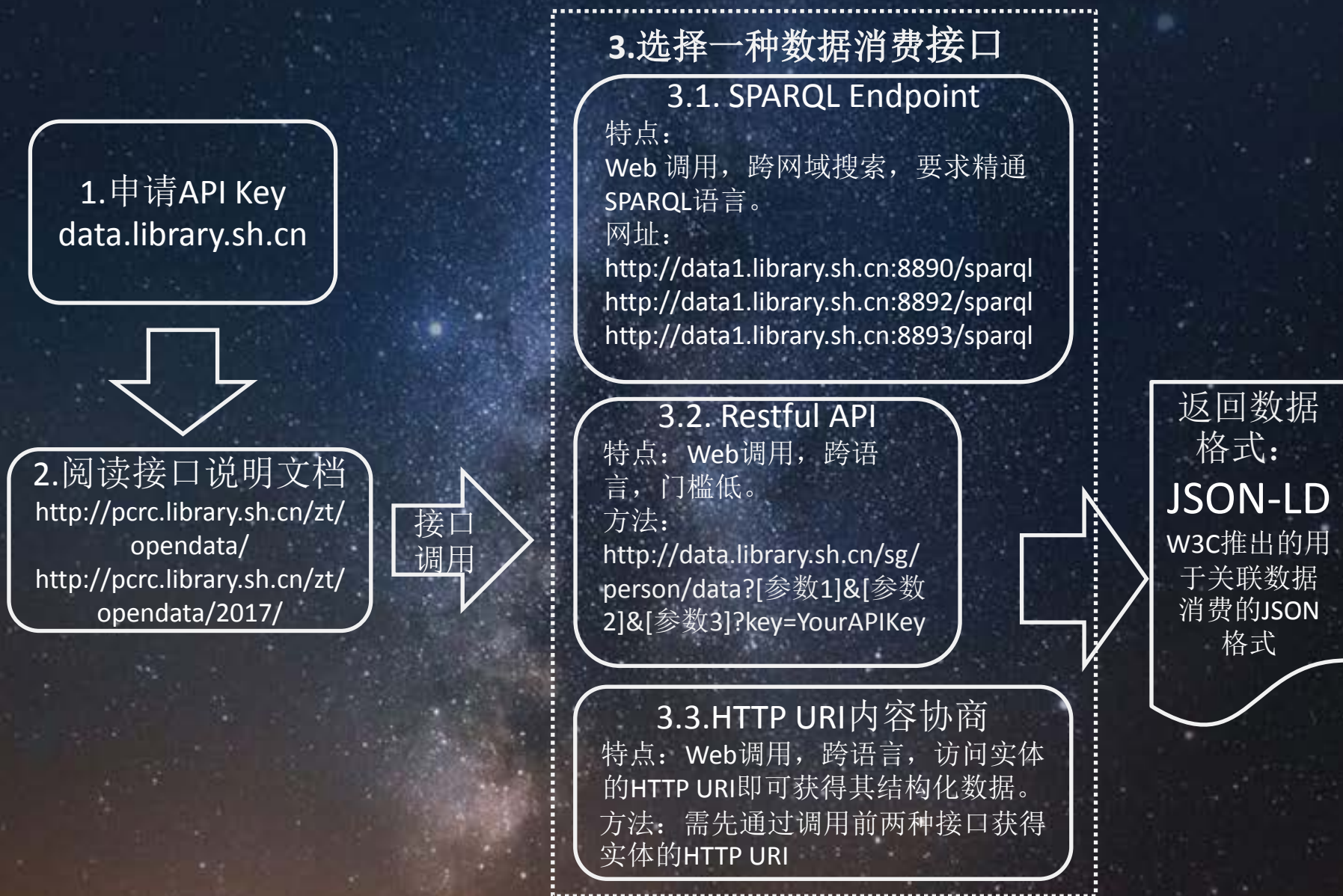
HTTP URI

JSON-LD

返回数据

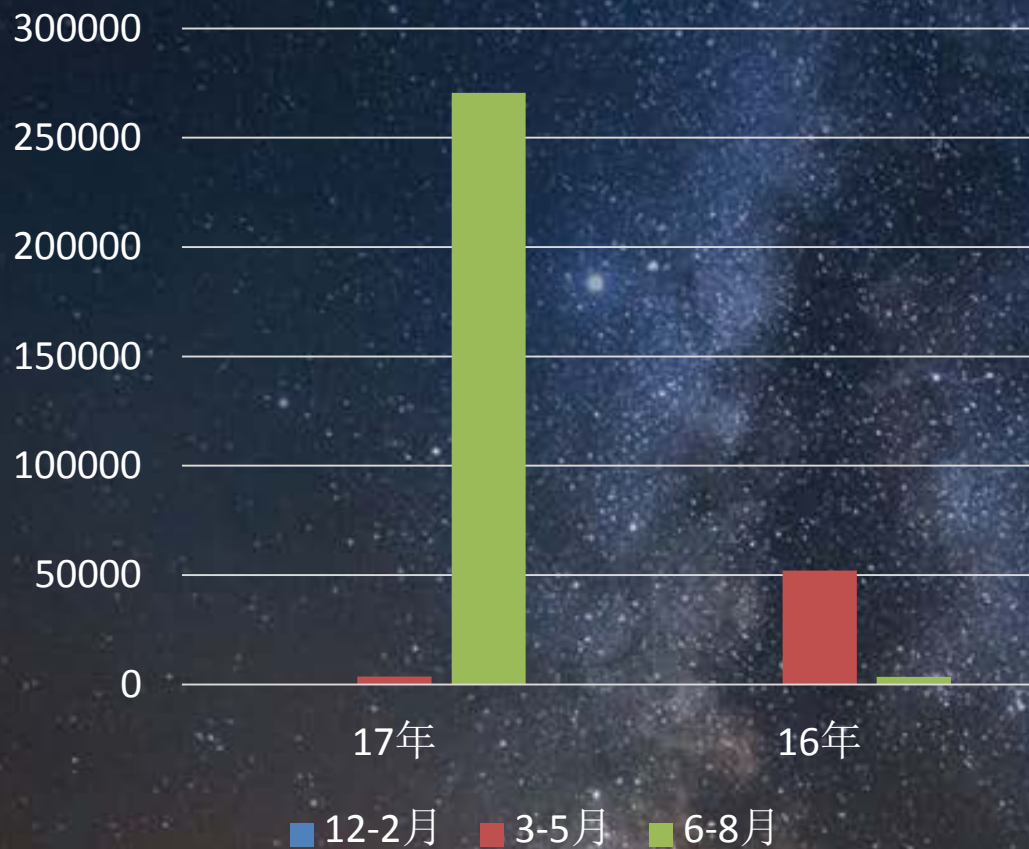
数据消费接口

开放数据接口调用方法及返回数据格式

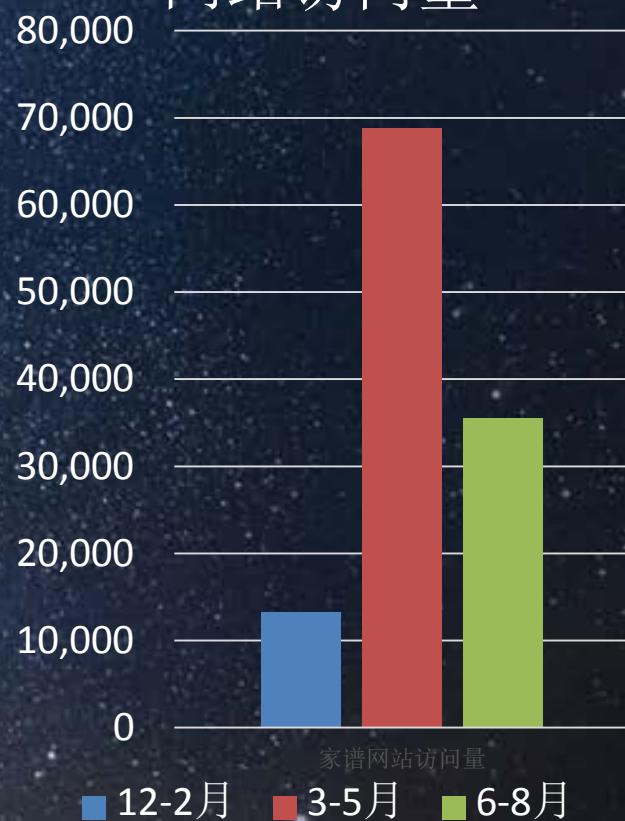


总结

API调用次数



网站访问量



后续工作

1. 数据扩充

- ◆ 人名规范库：79.7万
- ◆ 地理名词：1887
- ◆ 家规家训：307种228万字
- ◆ 古籍书目：516239条

2. 技术升级

- ◆ 框架改善
- ◆ 性能提升

一起来，更精彩！

Be together, Be better!

感谢各位的参与！

cjxia@libnet.sh.cn