

从数字图书馆到数字人文——

图书馆数字人文的背景、方法、技术和使命

From Digital Library to Digital Humanities: The
Background, Methods, technologies and Mission of Libraries

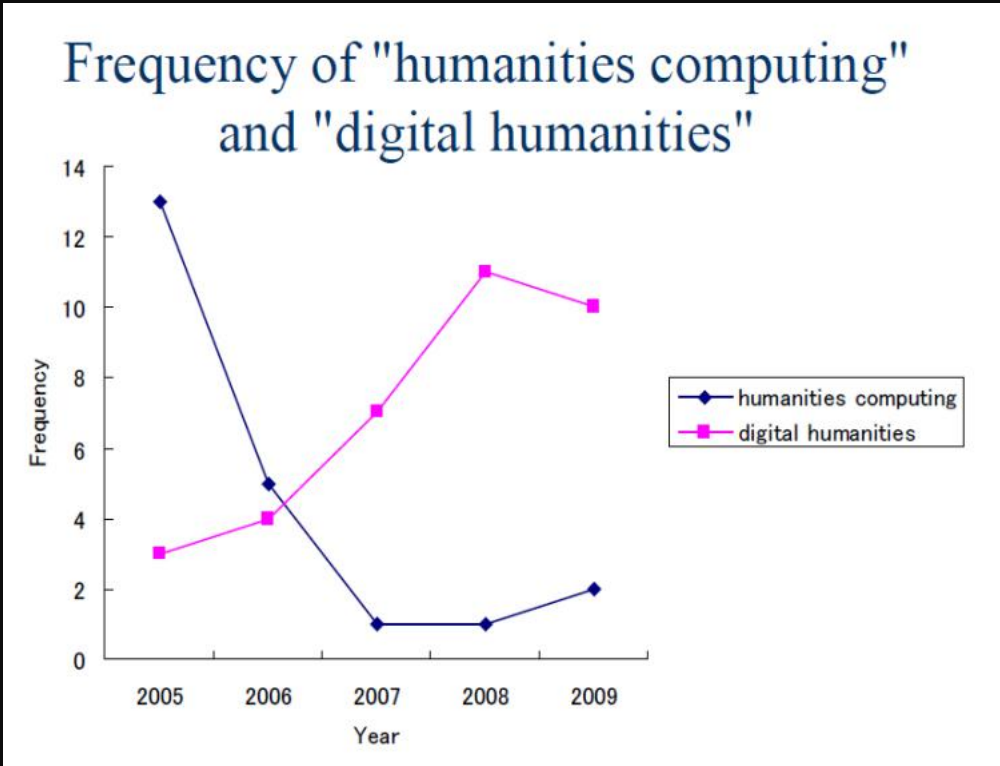
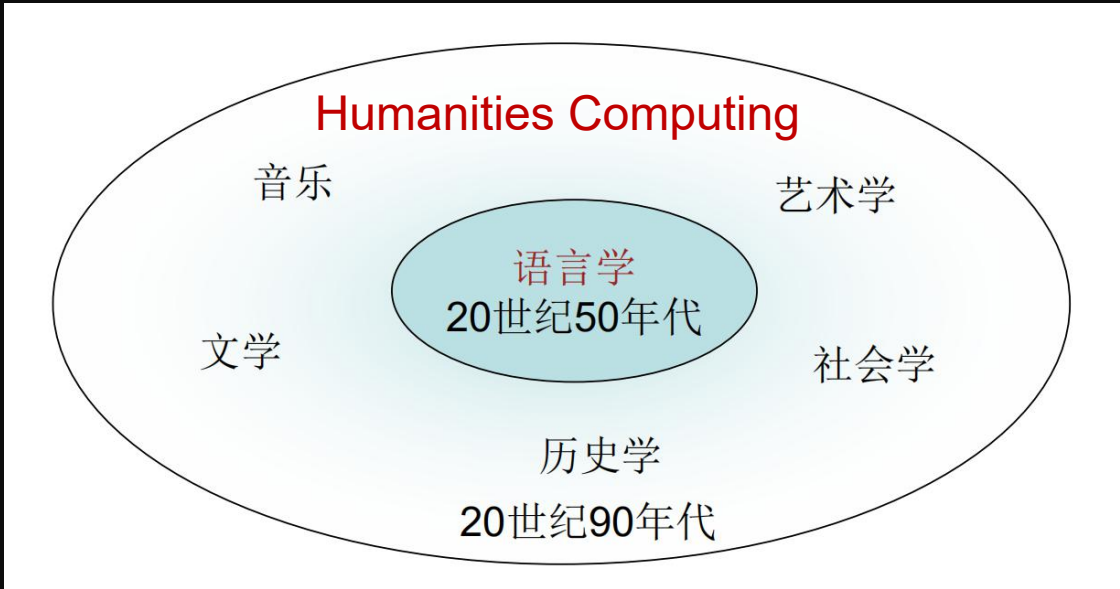
上海图书馆 夏翠娟
2018年6月 @****大学

DH 数字人文

Digital Humanities

现代数字技术+工具+方法：辅助人文科学研究

Application of digital resources and methods to humanistic inquiry. 应用数字资源与方法回答人文问题。 ——Waters 2013,4



Xiaoguang Wang, Mitsuyuki Inaba. Co-word Analysis of Research Topics in Digital Humanities (PP: 148-150), College Park, Maryland, USA: 2009.6

数字人文简史

1. 起源： 1949年意大利神父罗伯托·布萨（Roberto Busa, 1913-2011）开始使用电脑处理阿奎那（Thomas Aquinas）的全集，半自动地生成索引，开启了利用计算机处理机读文本，进行大规模文本分析计算的“人文计算”时代。

2. 发展： 2000年后，互联网的兴起，数据量大、生产加快、存取方便，不仅是计算，多种手段结合，门槛降低、普及面广，“人文计算”发展成为“数字人文”。

3. 勃兴： 2014年，Schich M, Song C, Ahn Y Y, et al. A network framework of cultural history[J]. Science, 2014, 345(6196):558-562. 数字人文改变了人文研究的范式，也促进了整个上下游（出版行业、文化记忆机构、学者）全面参与，带动了人文研究环境的革新。

欧美数字人文典型案例

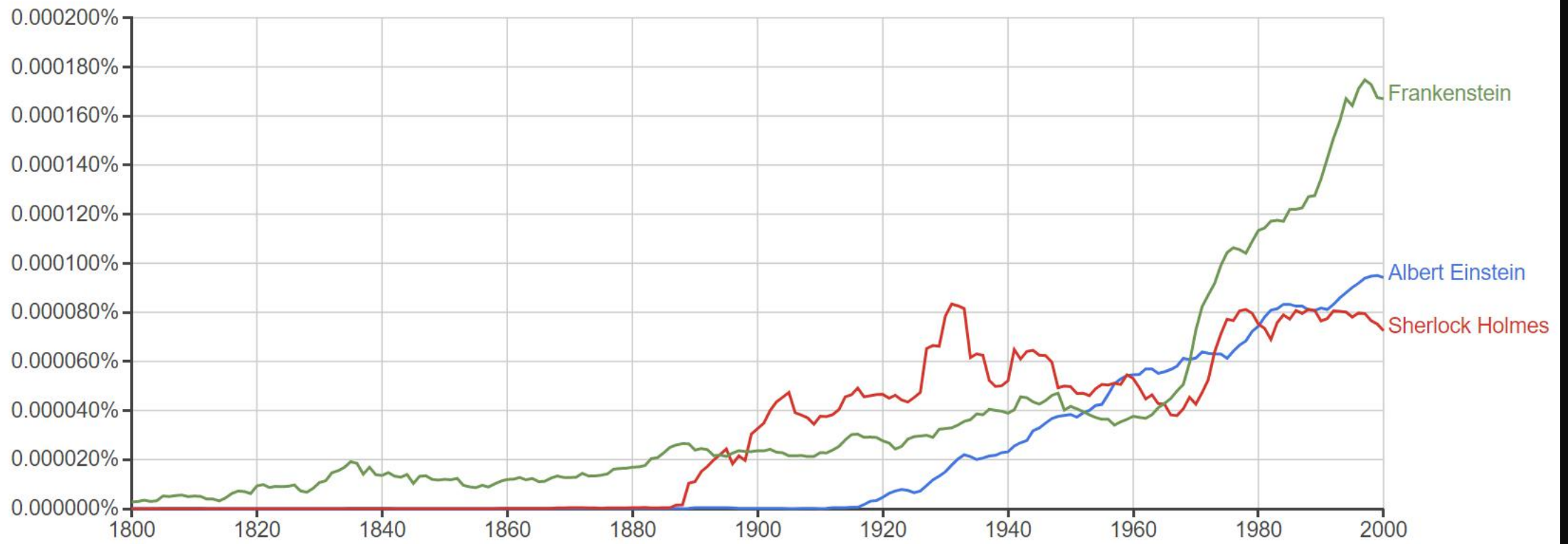
1. Google Books Ngram Viewer
2. Mapping the republic of letters
3. Venice Time Machine

Google Books Ngram Viewer

Graph these comma-separated phrases: ☐ case-insensitive

between and from the corpus with smoothing of .

[Search lots of books](#)



<https://books.google.com/ngrams>



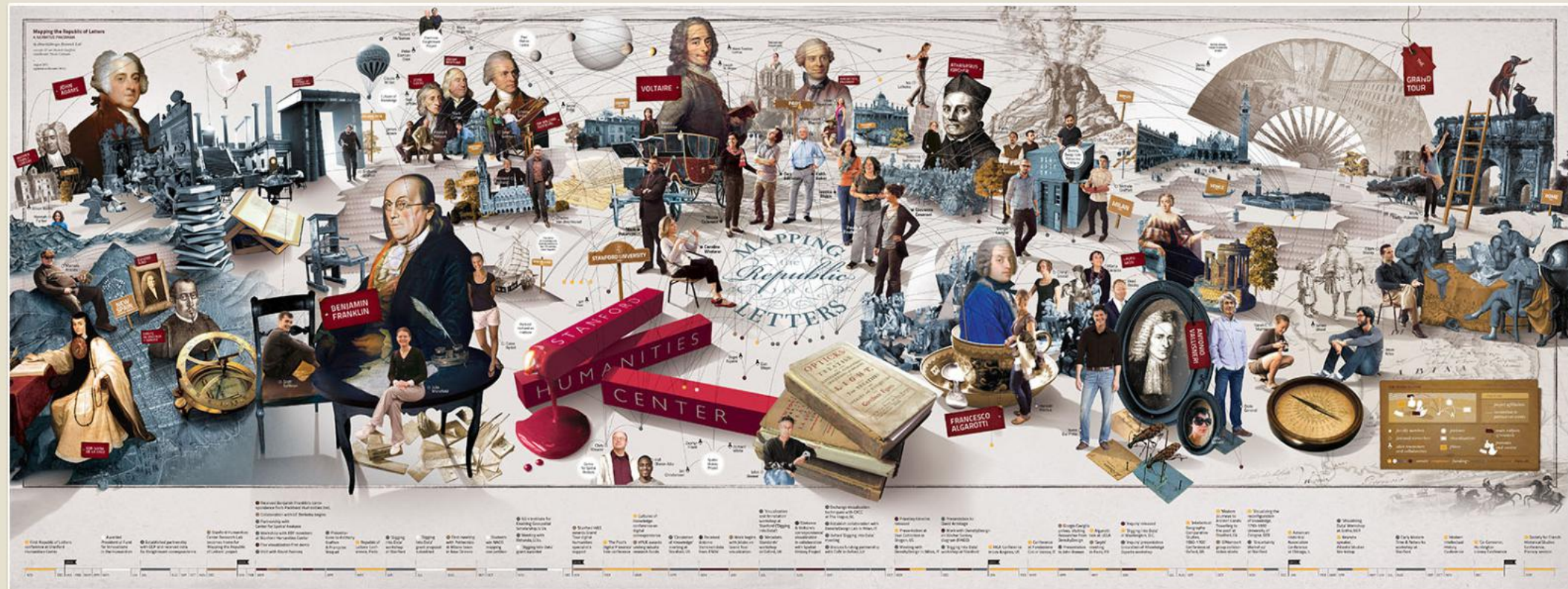
MAPPING THE REPUBLIC OF LETTERS

Home

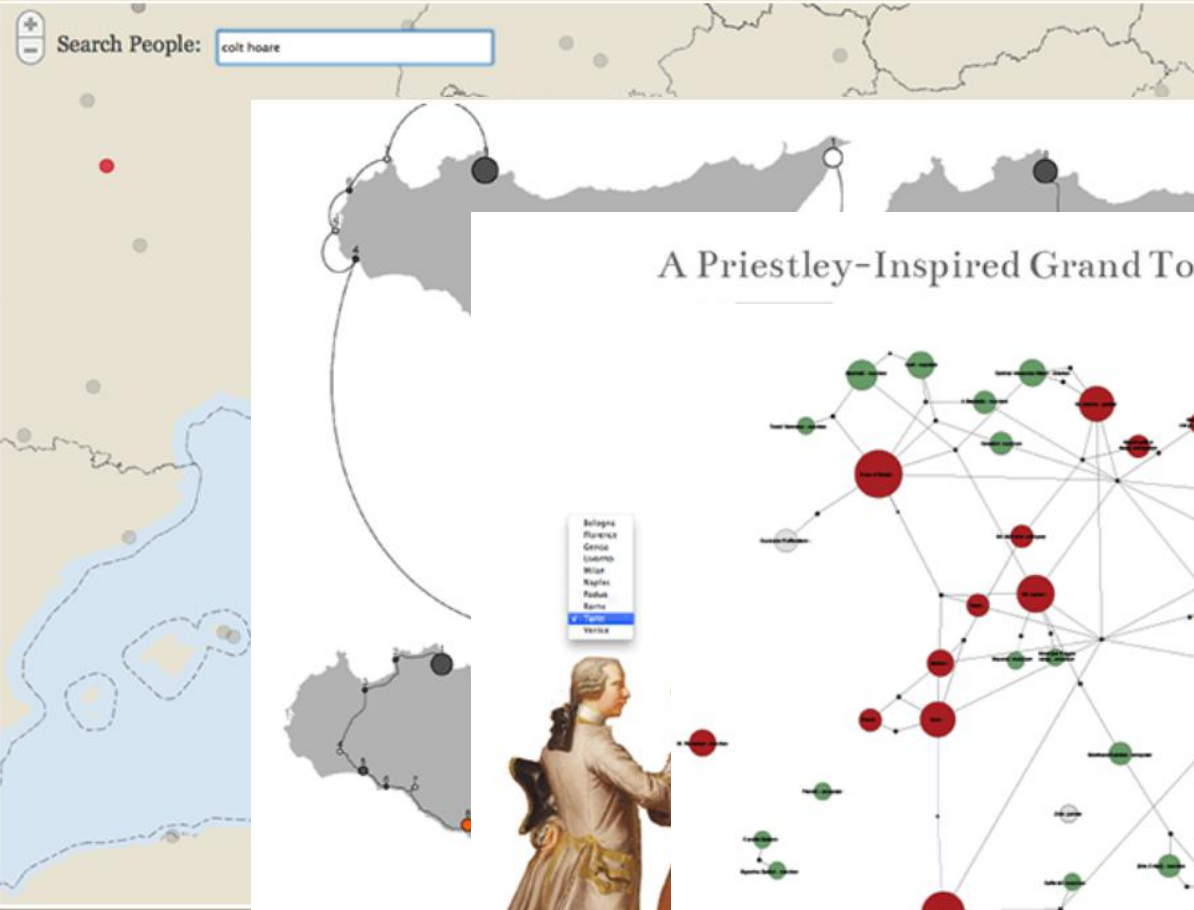
Case Studies

Publications

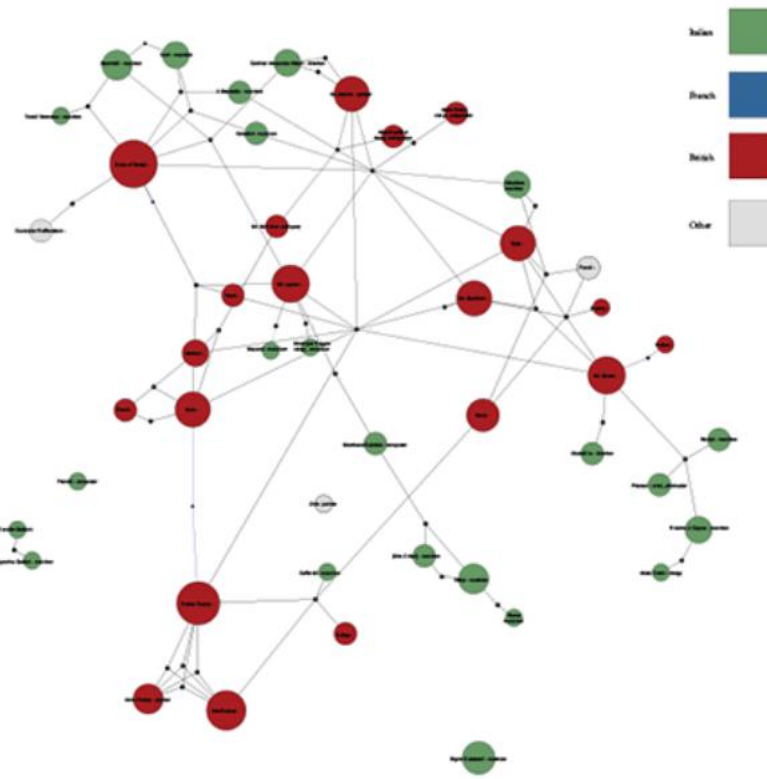
Contact



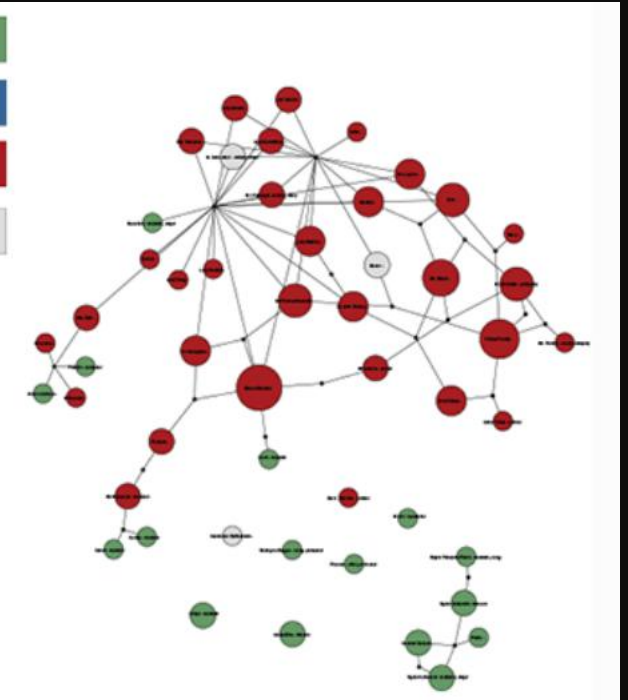
INFORMATION → DATA → VISUALIZATION



A Priestley-Inspired Grand Tour Timeline

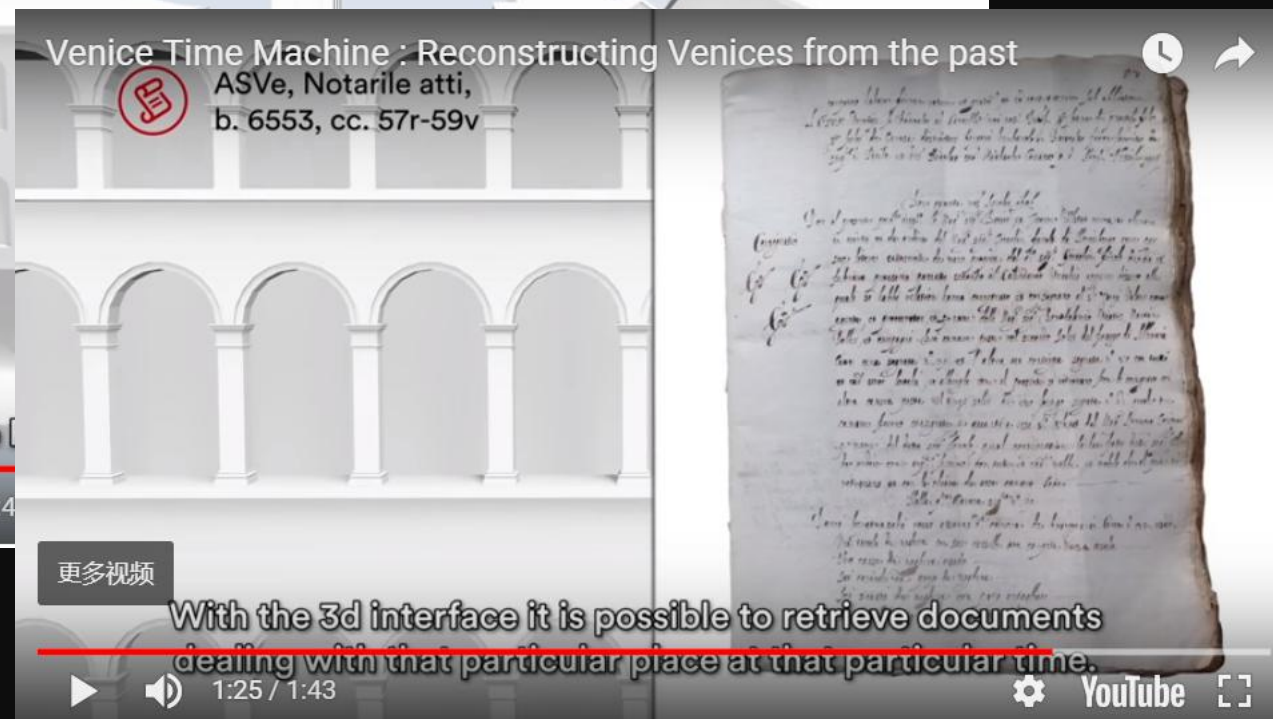
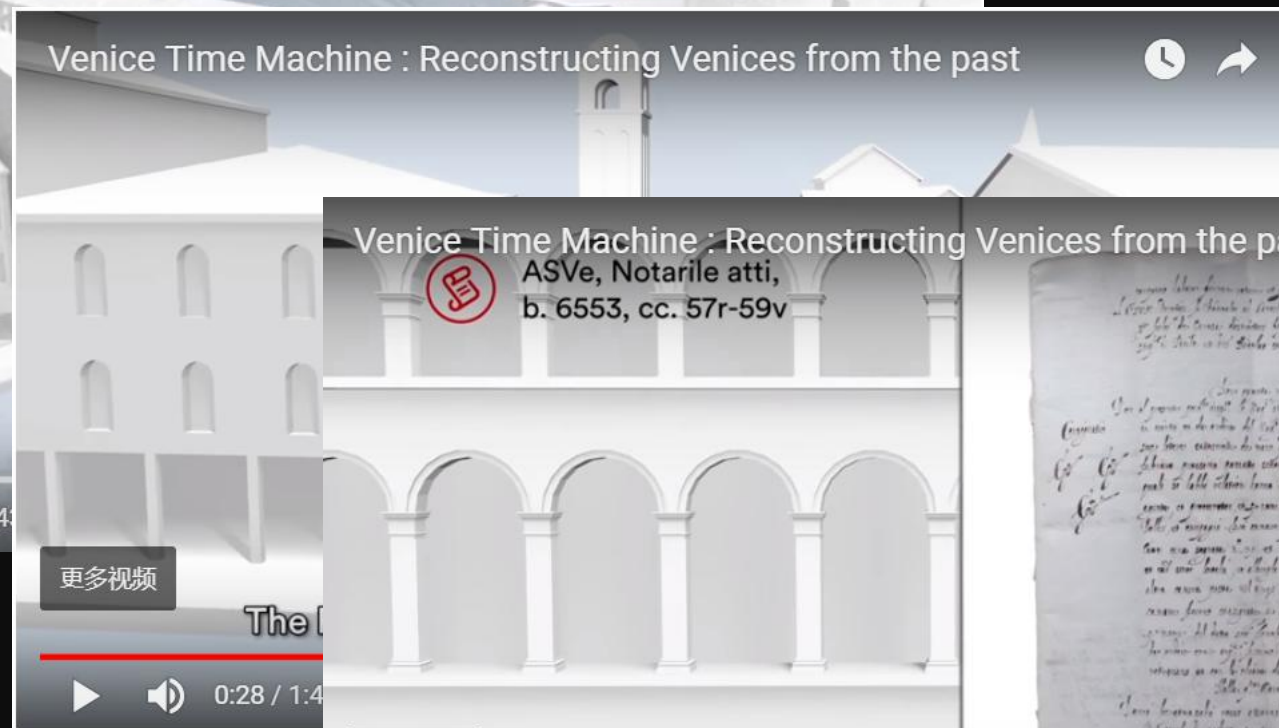
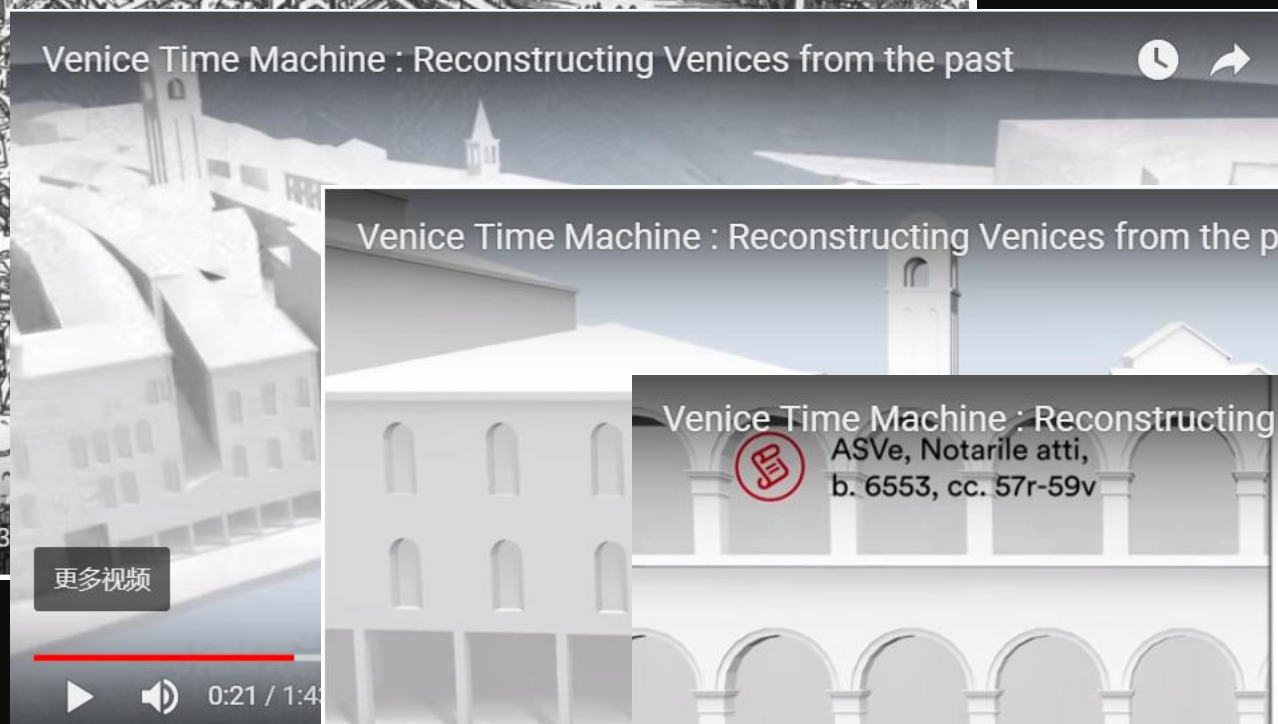


Rome, September, 20-October 14



Rome, November 10-22

<http://republicofletters.stanford.edu/casestudies/grandtour.html>

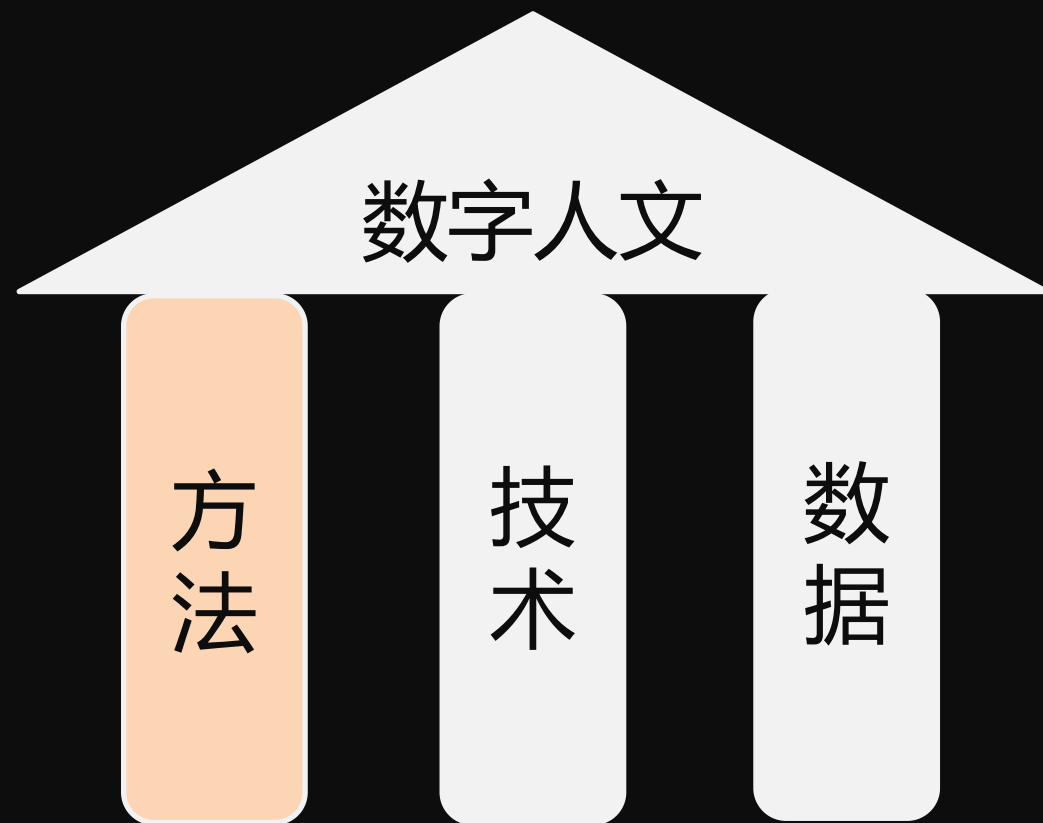


<https://vtm.epfl.ch/page-109840-en.html>

With the 3d interface it is possible to retrieve documents dealing with that particular place at that particular time.

方法：

词频统计，时空分析，社会
网络分析，实体关系分析，
循证实践，公众科学



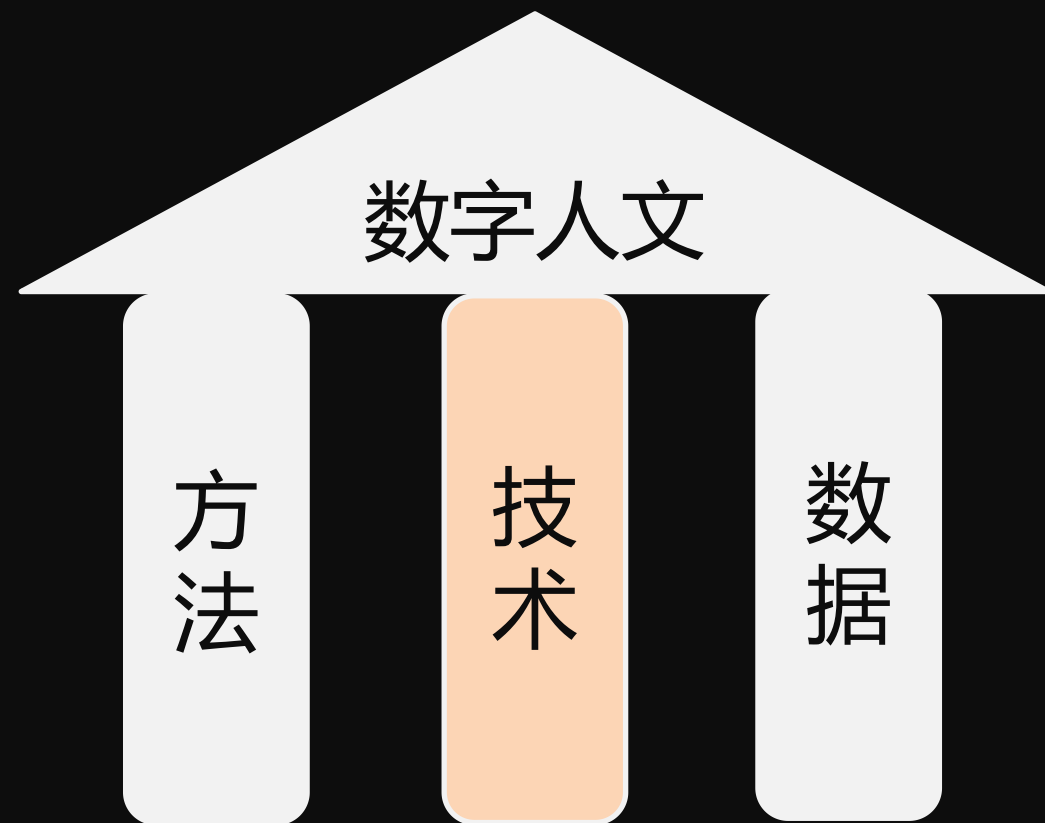
技术：

知识组织：语义网，关联数据，知识图谱...

基础通用：云计算，大数据，sns，微服务

分析计算：文本分析，NLP，NER，AI...

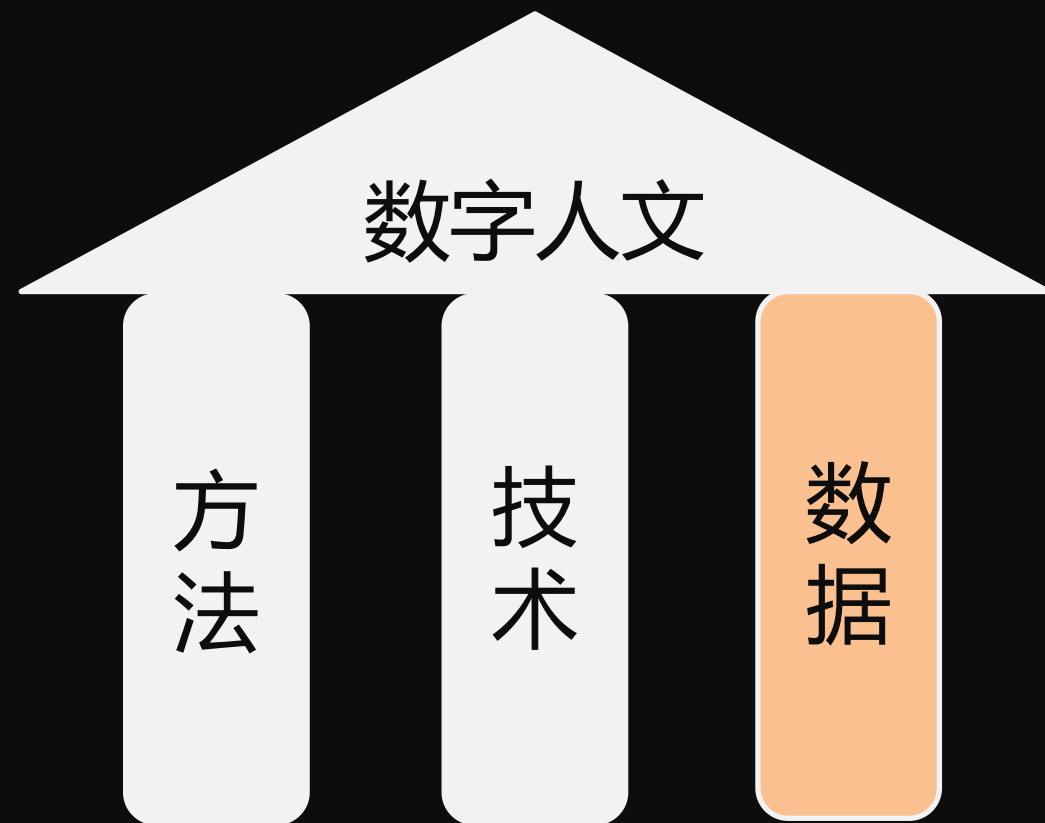
展示呈现：GIS，可视化...



数据：

Using the word "data" to mean "transmittable and storable computer information" was first done in 1946.

<https://en.wikipedia.org/wiki/Data>

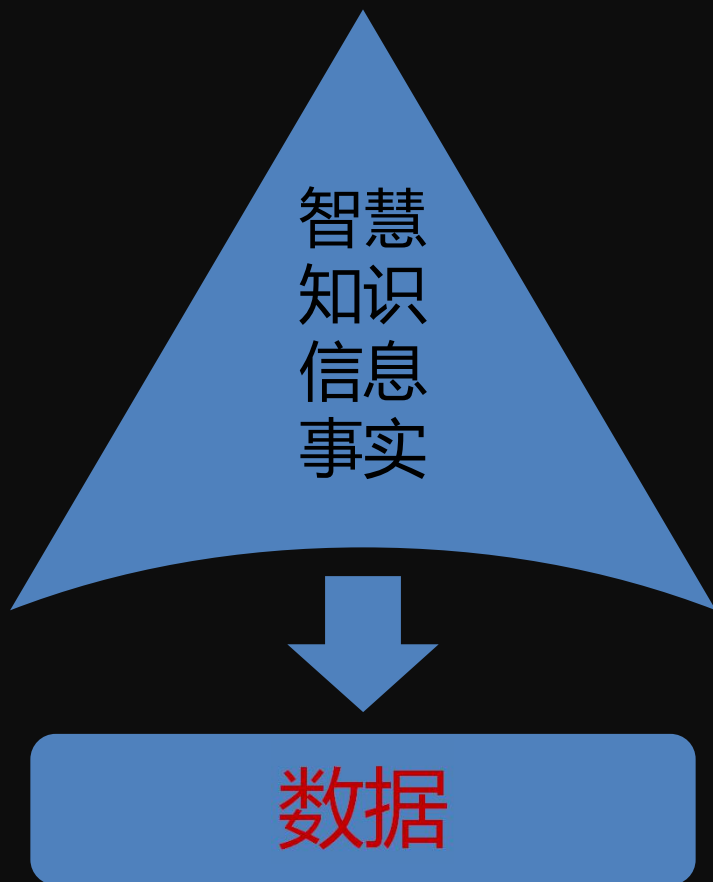


A blue triangle pointing upwards, containing a vertical stack of five Chinese characters. The characters are "智慧" (Wisdom), "知识" (Knowledge), "信息" (Information), "数据" (Data), and "事实" (Fact). The character "数据" (Data) is highlighted in red.

智慧
知识
信息
数据
事实

数据逐步成为新型生产力和能源

多媒体，富格式，
大规模，细粒度，
长时间，多维度，
可获得，可共享，
可计算，可分析，
可信，可预测，
可溯源，自解释



世界的本质是数据。

——维克托·迈尔-舍恩伯格

《大数据时代》

数据文化是尊重事实、强调精确、
推崇理性和逻辑的文化。

——涂子沛 《数据之巅》

传统人文研究的范式

1. 定性分析而非定量分析，小样本，个性化
2. 查阅、收集、比较、选择、分析、归纳、考证、注解、记忆
3. 思辨、演绎、推理、解释、叙述
4. 表达、呈现

数字人文研究的范式

1. 定量分析辅助并证明定性分析，大样本，全景式，公众参与，结论可重现
2. 查阅、收集、比较、选择、分析、归纳、考证、注解、记忆（机器辅助）
3. 思辨、演绎、推理、解释、叙述（用数据说话）
4. 表达、呈现（可视化技术）

汉学领域的数字人文进展

1. 哈佛大学社会计量学系CBDB，哈佛燕京图书馆
Ctext Project，德国马普所方志分析，Markus
2. 中研院CCTS、台湾大学Docusky、台湾师范大学
TBDB、法鼓文理學院佛学资料库、人物资料规范库
3. 武汉大学数字人文研究中心，复旦大学CHGIS，北
京大学，南京大学，上海图书馆

“数字图书馆(DL)” Vs. “数字人文(DH)”

数字图书馆

数字化加工

数字化形态

文献管理和保存

文献查阅服务

数字人文

内容分析知识组织

基于事实和逻辑的数据

共建共享和数据开放

基于科学方法的人文研究

数字人文是数字图书馆建设达到一定规模后的必然发展方向。

Digitizing the Humanities: A Future for Libraries. *Kathleen L. Sacco, etc.* 2015

图书馆可为数字人文贡献什么

1. 数字化资源? ——> 数据化资源!

从元数据记录中提取数据, 从数字化全文中挖掘数据, 从多媒体资料中识别数据, 从用户交互过程中获取数据...

2. 方法: 知识组织、书目控制、规范控制、文献循证

3. 平台: 数据基础设施

融合的数据集: 馆内融合, 馆际融合, 跨界融合, 跨网域融合

开放的数据服务: 基于互联网、多途径、跨平台、跨领域

易用可用的工具、软件: 培训、教育、环境

图书馆面对数字人文的使命和责任

1. 从数字化走向数据化，充分挖掘数据的价值；
2. 支持以数据驱动研究、学习和交流，革新图书馆学情报学的方法论体系；
3. 推动数据开放、合作、共享，让数据在流动中增值；
4. 参与数据基础设施建设，构建数字人文研究环境；
5. 建立系统性的数据保存、监护、分析、展示的标准规范体系，保证人文研究结论的可靠性和可验证性；
6. 开展数字人文方法、技术、工具的教育和培训，帮助传统人文学者熟悉数字人文研究范式.....

上图数字人文项目建设——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>



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

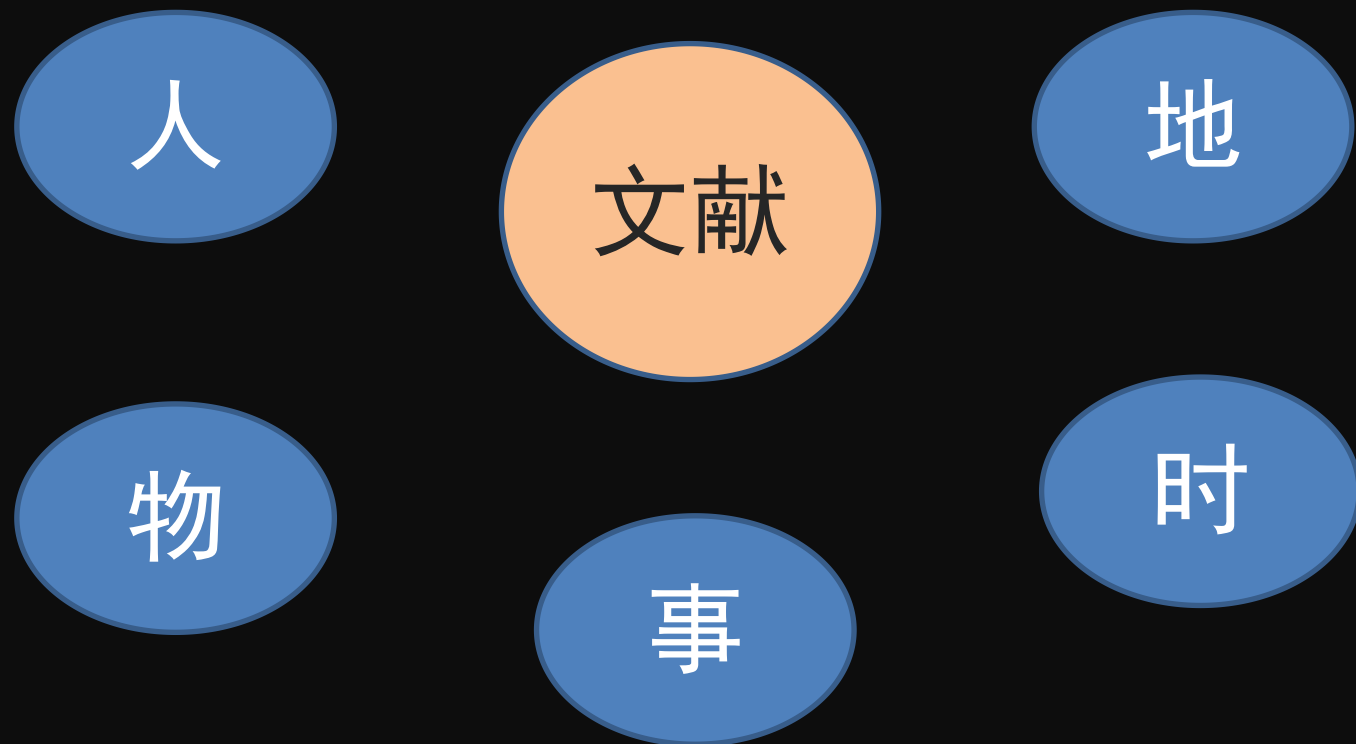
建筑

事件

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

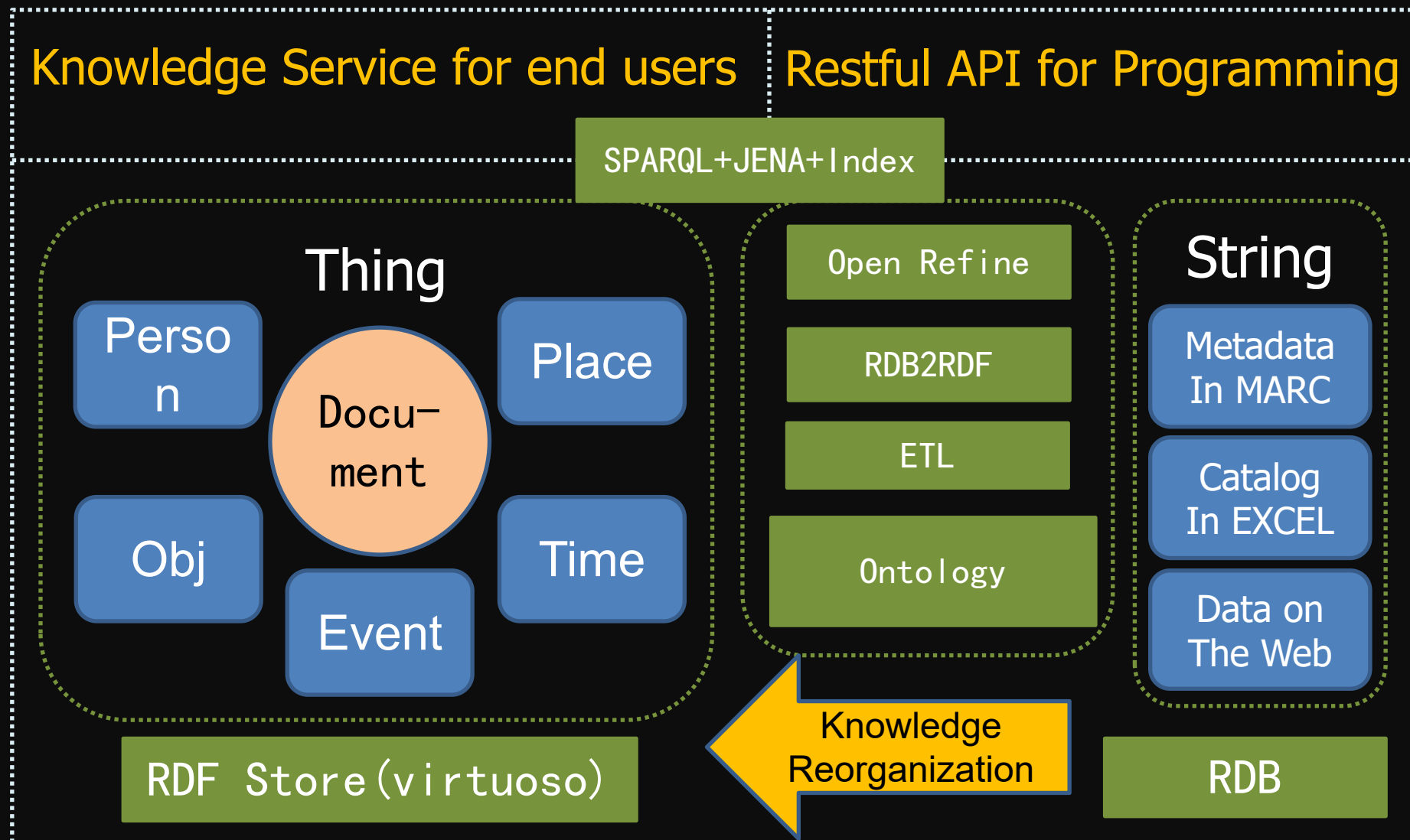
上图数字人文平台——方法

String to Thing (从字符串到对象)

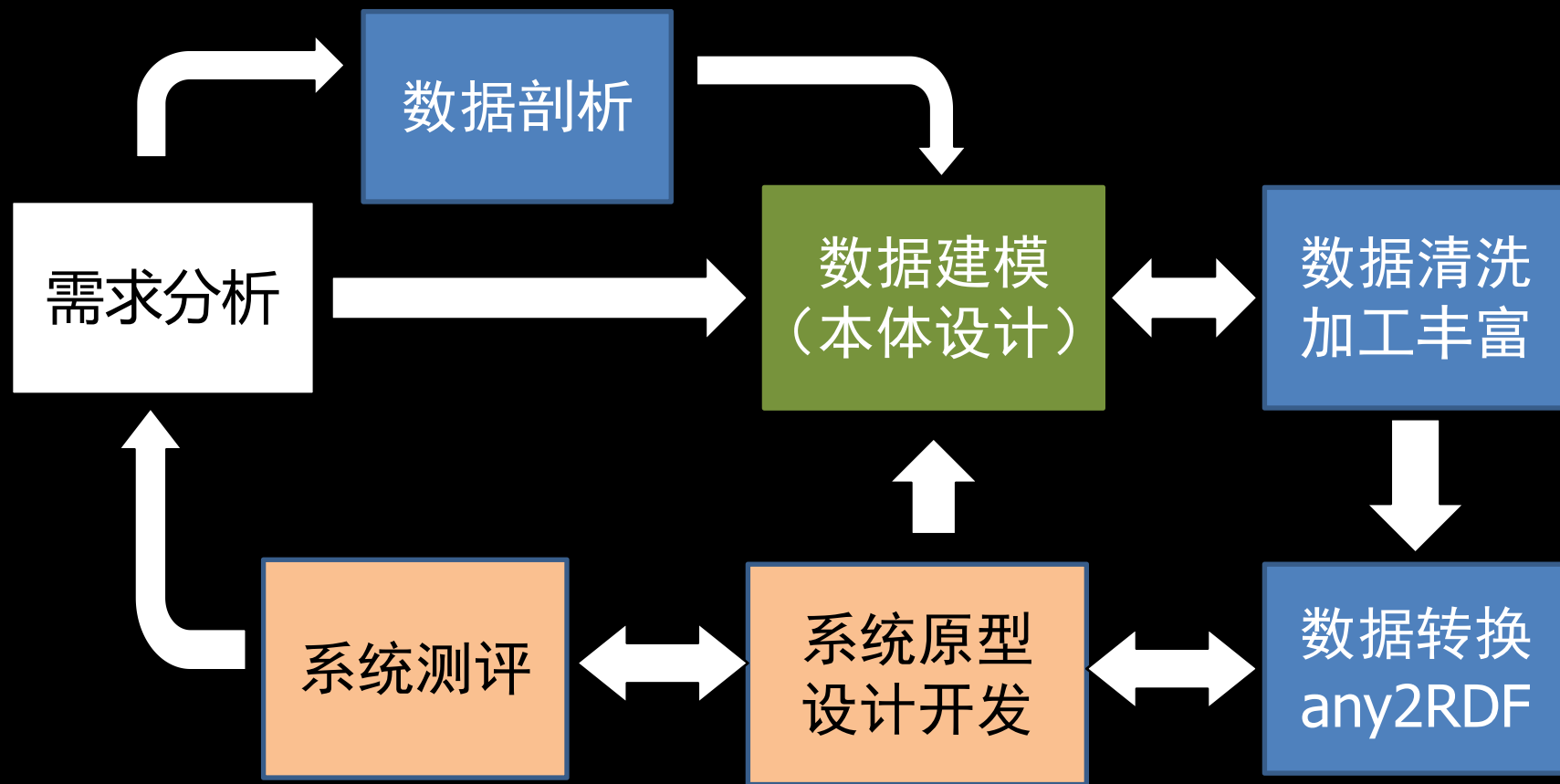


建设人地时事物基本知识库和文献知识库
成为面向人文研究的数据基础设施的一部分

上图数字人文平台——技术框架



上图数字人文平台——建设流程



上图数字人文平台——数据开放

data.library.sh.cn

上海图书馆
上海科学技术情报研究所
上海图书馆开放数据平台

本网站是上海图书馆的开放数据平台，将陆续以关联开放数据（Linked Open Data）的方式向互联网公开发布各种术语表、规范档、书目数据以及各种数据清洗和转换工具等。提供各种数据消费接口供开发人员调用，以促进数据的开发获取、共享和重用。调用接口前，请注册API Key。

人名规范库

More...

华人姓氏列表

More...

中国历史纪年表

More...

地理名词表

More...

机构名录

More...

武康路

More...

家谱本体

More...

手稿及档案本体

More...

古籍本体

More...

Copyright©上海图书馆 2015 | 非特别注明，本站遵循cc2.0协议（署名-非商业性使用-相同方式共享）

开放数据应用开发 启动 竞赛公告2018.4.1



扫一扫，进入参赛主页

上海图书馆 | 全球征集

官方网站 URL:
<http://opendata.library.sh.cn/>

上海图书馆开放数据应用开发竞赛已成功举办两年，为推动我国图书馆界开放数据的发展做出了有益探索。今年，我们将与多家机构合作，打开数据的宝库，再次聚焦历史人文数据，面向全社会征集以开放数据为基础的优异移动应用产品原型或服务创意，以期更加充分地释放开放数据的价值，最大程度挖掘其背后的应用潜力。

更加丰富的历史人文开放数据
和10万元大奖等着你！

一起来更精彩



4.1-8.15
等你来呈现
最完美的作品



此次竞赛项目征集及通知将在竞赛官方网站及文化部公共文化研究上海图书馆基地官方微博/微信上，请各位参赛者及时关注。

电话：021-64281593（总机） 邮箱：cylh@libnet.sh.cn

主办单位：上海图书馆（上海科学技术情报研究所） 文化部公共文化研究上海图书馆基地
协办单位：CADAL项目管理中心 上海市计算机学会 复旦大学图书馆 广州图书馆有限公司
支持单位：国图集团有限公司



Our DH team @ shi

谢谢!

xtykc@yeah.net