# 数字人文：图书馆实践的新方向

2017年6月·贵州

北京大学　朱本军

bjzhu@pku.edu.cn

# 汇报提纲

1  “数字人文”的现状

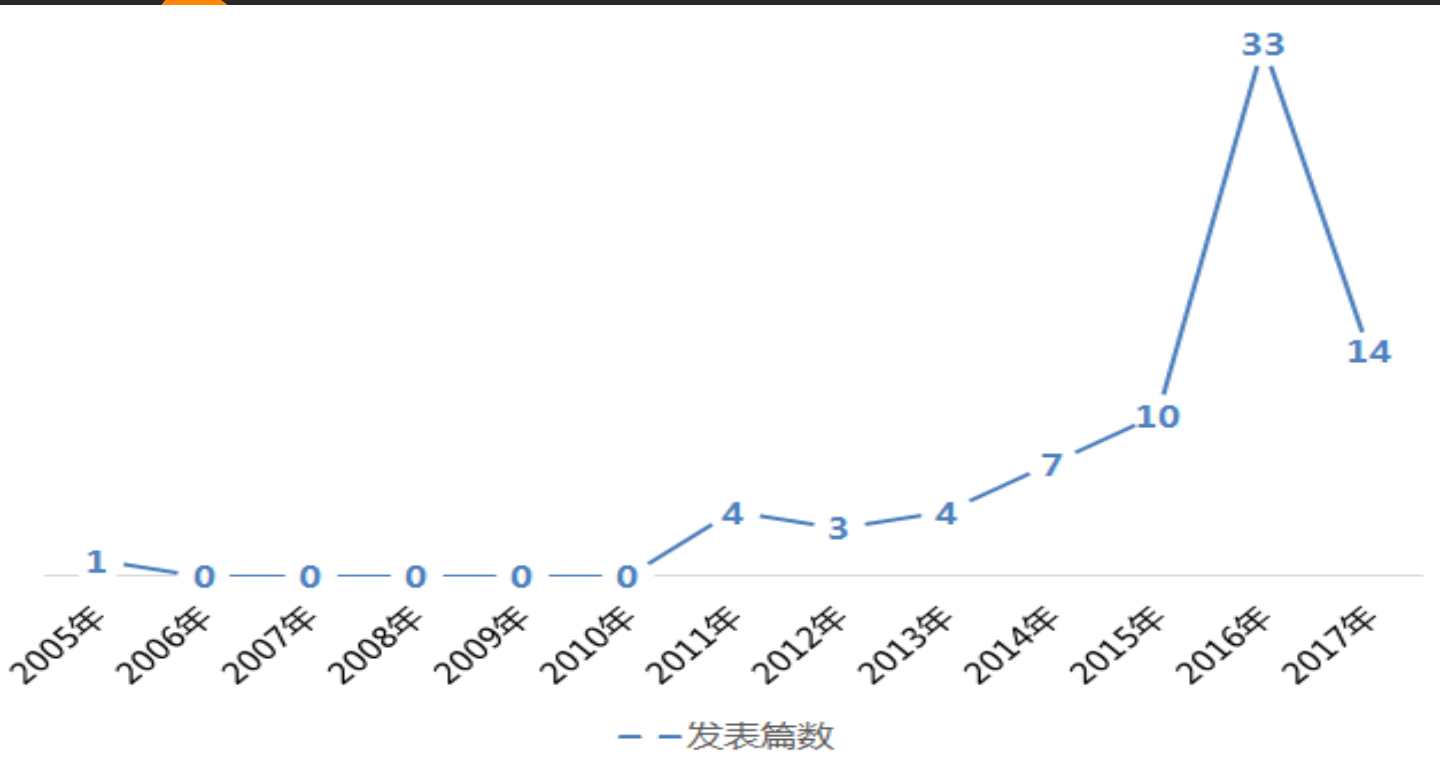
2  数字人文的内涵

3  图书馆实践的方向

4  北京大学数字人文做法

# 数字人文的现状

# 1、中国大陆数字人文刚刚起步



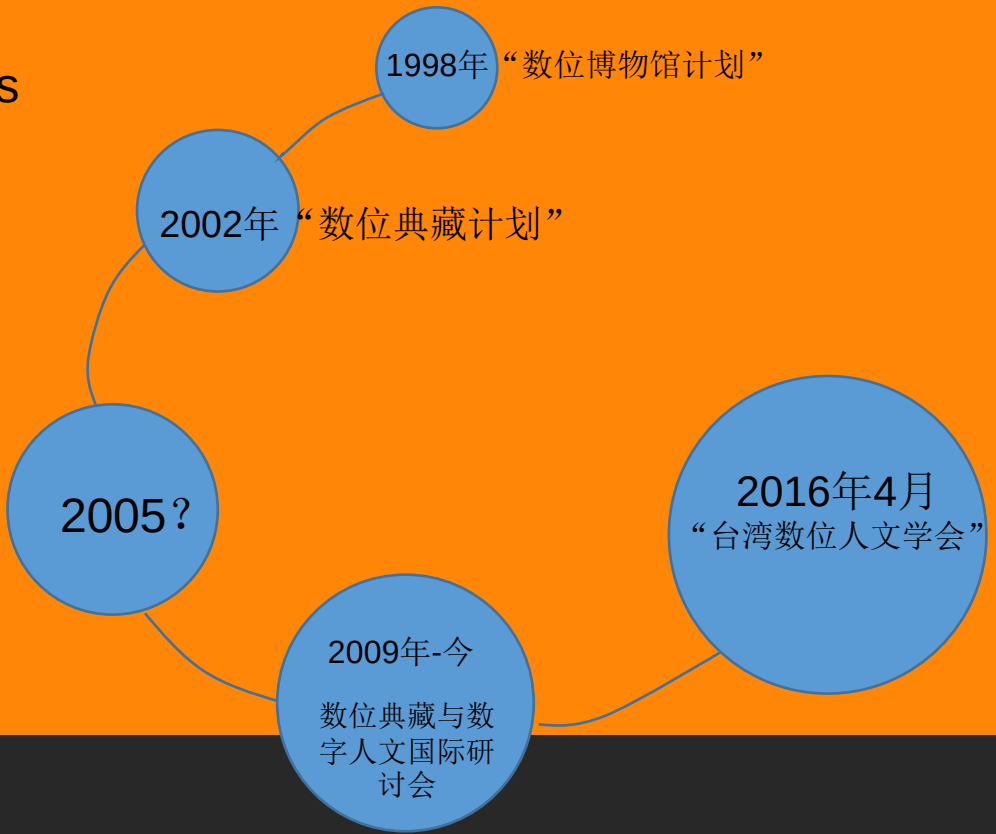

CNKI直接以“数字人文”、“人文计算”为标题发文趋势

# 1、中国大陆数字人文刚刚起步

- 数字人文机构或团队建设
  - 武汉大学数字人文中心（2011年）
  - 北京大学数字人文小组（2016年）
  - 南京大学数字人文小组（2017年）
- 数字人文相关学术活动
  - 学术会议
    - 2014年6月，上海图书馆“数字人文与语义技术”
    - 2015年12月，北、清、台“数字人文新动向——中国历代人物传记资料数据库暨Digging into Data工作坊”
    - 2016年5月，“北京大学数字人文论坛”（首届）
    - 2016年5月，“数字人文与清史研究”
    - 2017年5月，“北京大学数字人文论坛”（第二届）
    - 2017年7月，南京大学“数字人文:大数据时代学术前沿与探索”
  - 工作坊
    - 2016年，南京大学历史学院王涛副教授“数字工具与世界史研究”课程
    - 2017年3月，哈佛大学访问学者徐力恒博士在北京大学开设“数字人文研究技能与方法”读书会
    - 2017年4月，北京大学图书馆数字人文工作坊（第1期 SNA；第2期 GIS）

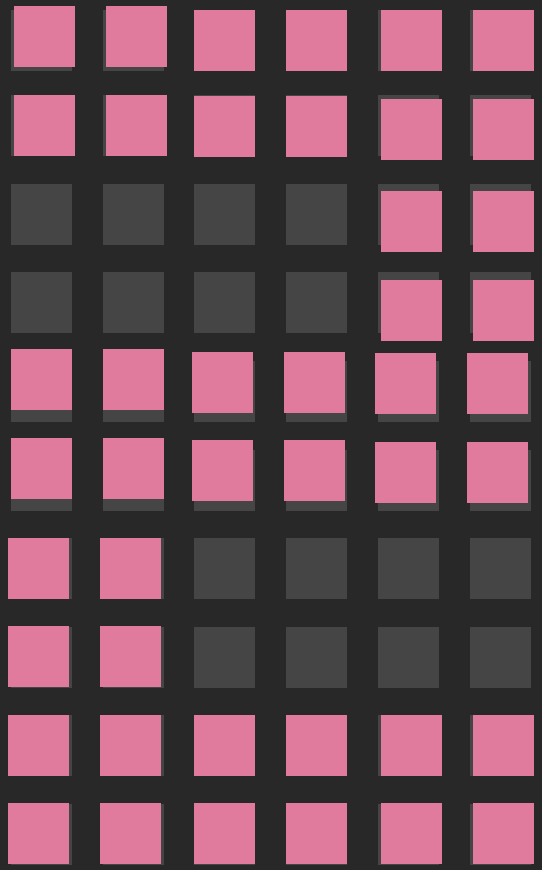
# 2、中国港澳台地区数字人文起步不久

Digital Humanities
||
数位人文



# 3、欧美高校数字人文羽翼丰满

- 超过183个直接冠以“数字人文”的中心、项目、实验室、团队或圈子
  - 宾夕法尼亚大学“Price实验室”、耶鲁大学“数字人文实验室 (DHLab)”
  - 斯坦福大学、牛津大学“DH计划”
  - 普林斯顿大学图书馆数字人文中心
- 国际学术会议
  - *Digital Humanities conference*
- 国际刊物
  - *Literacy and Linguistic Computing*(文学和语言计算)
  - *Text Technology*(文本技术)
  - *Computers in the Humanities Working Papers*(人文领域计算机应用工作论文)
  - *Digital Humanities Quarterly*(数字人文季刊)
  - *Companion to Digital Humanities*(数字人文指南)。
- 国际组织
  - *The Alliance of Digital Humanities Organizations*(ADHO,国际数字人文组织联盟)



# 数字人文的概念与内涵

# 1、数字人文的概念

1. Illinois：*at the intersection of computing and the disciplines of the humanities*
2. Enago Academy：*a simple re-branding of the old “humanities computing” field*
3. Wikipedia: *it is methodological by nature and interdisciplinary in scope. It involves investigation, analysis, synthesis and presentation of information in electronic form. It studies how these media affect the disciplines in which they are used, and what these disciplines have to contribute to our knowledge of computing.*
4. Steven Hayes：*Modelling and recording traditional humanities data sets in such a way that they can be read by both humans and machines.*
5. ……

# 1、数字人文的概念

Digital Humanities称谓的来源：

- 2001年，《Companion to Humanities Computing》改名为《Companion to Digital Humanities》

“humanities computing”？ “digitized humanities”？

“digital humanities”？

- 2005年，文学与语言计算协会（The Association for Literary and Linguistic Computing）、计算机与人文协会（The Association for Computers and the Humanities）为一个新的实体组织命名

“eHumanities”？

“humanities computing”？

“digital humanities”？

国际数字人文组织联盟（The Alliance of Digital Humanities Organizations，ADHO）

# 1、数字人文的概念与内涵

- 尽管从文字上界定比较困难，但是可以看到数字人文的努力方向是将数字技术与人文领域相结合，运用数字技术提出、探索和解决人文领域的各种人文问题，而不是人文领域的技术问题。

- 是否是数字人文的标准：
  - 不要看各种五花八门的技术、数据库
  - 主要看数字技术是否对“人文”（文、史、哲、艺术、宗教、语言）研究教学有促进？

# 2、数字人文的内容和努力方向（基于DH项目的总结）

## （1）节省传统人文学者在查找资料、文本处理、计算等方面低水平重复的时间精力

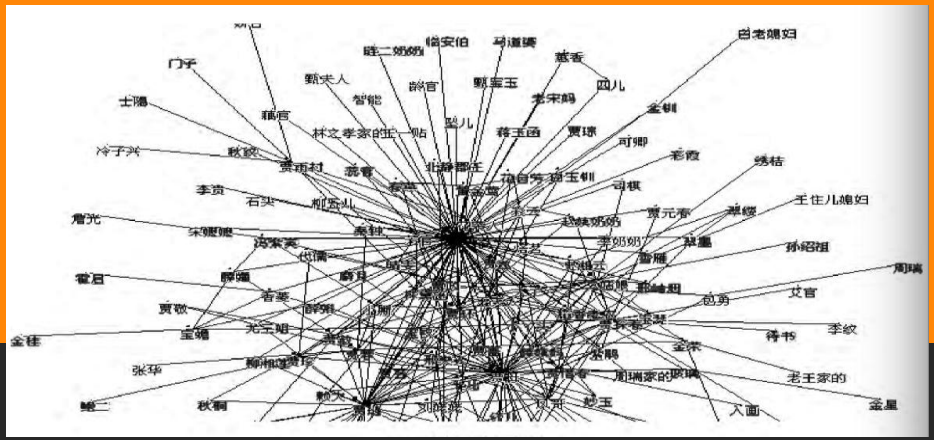
- **数字化（digitization）**，将非数字的人文内容加工转化为数字内容。
  - 人文学者不必在搜集资料上花费太多的时间精力
  - 典型的例子：JSTOR、中华经典古籍库、Google Books……

- **数据集（dataset）建设**，对非结构化的数字文本内容按照某种使用目的进行结构化标注，变成专门的数据集。
  - 切入到研究者的研究过程中，解决研究过程中的棘手问题
  - 典型例子：哈佛大学与北大联合的CBDB、台湾中研院地名规范检索

# 2、数字人文的内容和努力方向（基于DH项目的总结）

## （2）引入新的工具、方法解决传统的人文问题

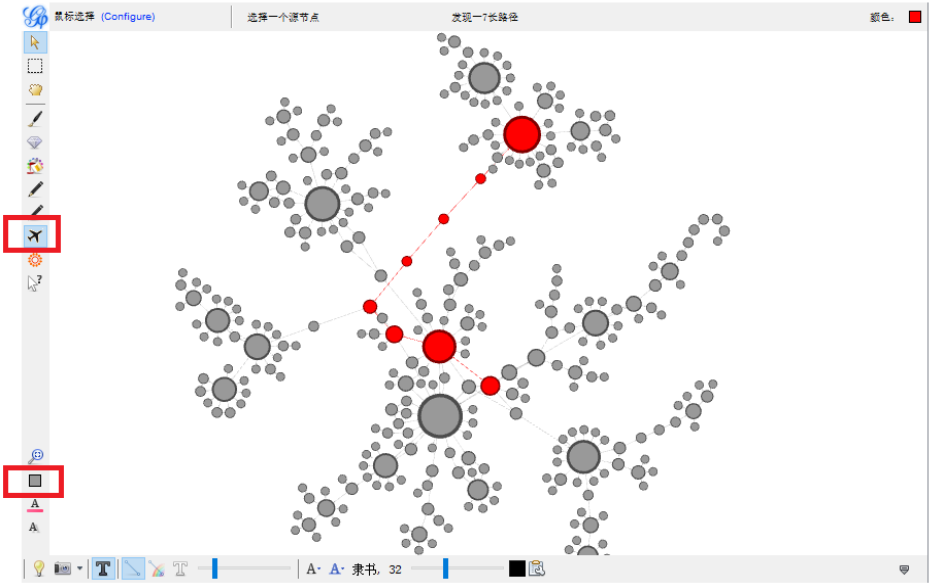
- **例子1：《红楼梦》的人物关系和阶级关系**
  - 乔先之. 论《红楼梦》人物的阶级关系[J]. 西北师大学报(社会科学版), 1974,(03):139-153
  - 陈蕾,胡亦旻,艾苇,胡俊峰. 《红楼梦》中社会权势关系的提取及网络构建[J]. 中文信息学报,2015,(05):185-193+203



# 2、数字人文的内容和努力方向（基于DH项目的总结）

## （2）引入新的工具、方法解决传统的人文问题

– 例子2：北宋的苏洵和南宋的王淮是什么关系？




蘇洵（第2个儿子->）蘇轍（第1个儿子->）蘇遲（儿子->）蘇簡（第1个儿子->）蘇謂（儿子->）蘇林（妻子的爸爸->）王師德（儿子->）王淮

# 2、数字人文的内容和努力方向（基于DH项目的总结）

## （3）利用数字思维方式或数字工具提出人文领域的新问题，并尝试回答新问题

- **例子1：敦煌莫高窟褪色壁画还原**
  - 不是简单的文字描述和推测，而是在对壁画颜料成分及艺术家经验知识的基础上，通过3D影像记录、色彩、图像处理、人工智能等知识还原原貌并模拟褪色过程

《我爱发明》 20160907 数字敦煌

- **例子2：河流改道或海岸线变迁**
  - 不是通过文字记载、地质沉积物等方式进行推理还原，而是直接通过大时间尺度的卫星遥感影像直接还原变迁的过程


←中国三峡水库蓄水前后

秘鲁Ucayali河流侵蚀河床
形成一个个牛轭湖→

# 2、数字人文的内容和努力方向（基于DH项目的总结）

（4）利用数字思维对传统人文领域进行创造性破坏和建设

- **例子：全息《清明上河图》**


# 图书馆实践数字人文的方向

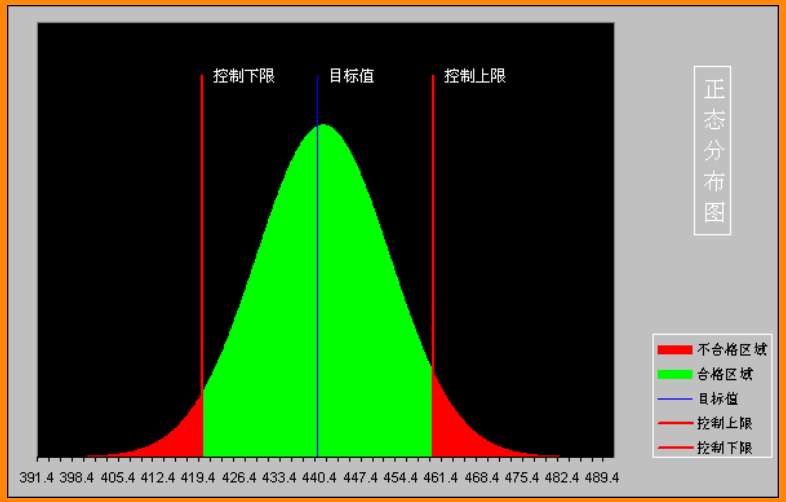
# 1、实践应该注意避免的误区一

## 误区一：数字人文=数字文科/数字人文社科

### 人文学科与社会科学的差别

- 社科学者点评人文（“赤壁之战”）
  - 整部《三国演义》所告诉大家的就是我们中国人的劣根性，什么劣根性？中华文化的“精髓”之一是什么？投机取巧，**它不但讲投机取巧，甚至还特别强调小概率事件**。我举个例子，你们知道赤壁之战要告诉大家什么吗？诸葛亮同志在那边装神弄鬼，借东风。赤壁之战借东风是胜败的关键，如果风不转向的话，无法进行火攻。因为当时是冬天，是北风，所以一定要有东风才行，所以那个同志就忽悠老天爷。运气太好了，东风终于来了，诸葛亮同志牛啊，结果用火攻打败了曹军。你可以考证是不是确有其事，这不重要。你们有没有思考过这个问题，诸葛亮同志如果借不来东风怎么办？百万将士的生命将置于何处？一个为政者竟然敢相信诸葛亮先生借东风，这本身就是不合理的，上百万人的性命就寄托在这么一个小概率事件上，这不是投机取巧是什么，而且我们中国文化居然这么强调这件事，甚至还拍成电影，大加赞扬，什么结果？这会使我们每一个中国人都认为投机取巧是对的，它是合情合理的。有没有听过四两拨千斤啊，你们首先问问自己，四两拨千斤拨不了怎么办，不是被千斤压死了嘛。中华文化为什么不能强调千斤拨四两呢，有必胜的把握为什么不好？从《三国演义》就能看得出来，中华文化是崇拜小概率事件的，是非常危险的。——郎咸平浙大演讲
- 人文学者下社科结论
  - 黄文仪《浅述汉代五铢钱制度建立的意义》（人文学者总结社会科学）：铜币铸造权收归中央,币制的统一,则使西汉初年以来长期存在的币值不稳,货币流通紊乱的问题获得解决,并建立起五铢钱制度。这一统一健全的货币制度的建立,又转而促进了社会经济的发展,以及封建中央集权国家的统一和巩固。

# 研究范式不一样，产生了巨大的分野




**社会科学学者：**人文只能称为学科，不能称为科学，科学是可以重复的，要靠数据说话（定量分析）

**人文学者：**落脚在文、史、哲，及其衍生出来的艺术、美学、宗教/伦理等；

不要将“人文”“社科”混为一谈

- 人文学者的研究范式
  - 定性研究（Qualitative research）或小样本定量
- 人文学者的创作
  - 基于基础材料/信息的思辨（Critical thought）、演绎推理（Deductive Reasoning）、解释（Hermeneutic）、叙述（narration）

**特点：个体性非常强、思维无固定的套路**

There are a thousand Hamlets in a thousand people's eyes.
一千个人眼里有一千个哈姆雷特

# 数字人文：对人文研究范式的改变

- **从定性→定量辅助+定性，将社会科学领域的一些研究方法合理地引入人文领域（量化史学）**

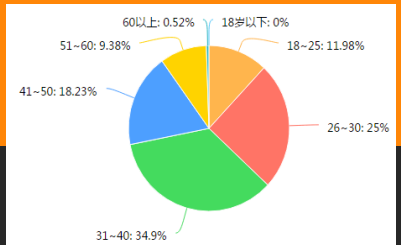
- **数字技术介入到**传统人文学者教学科研中的查找/探索（discovering）,注解/标注（annotating）, 对比/比对（comparing）, 取样（sampling）, 阐释（illustrating）, 表达/呈现（representing） 等活动节点

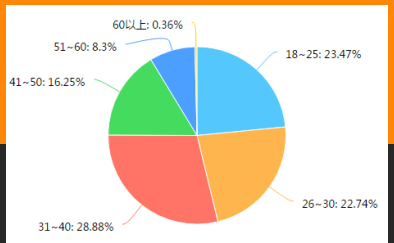
# 1、实践应该注意避免的误区二

## 误区二：过分强调华而不实的技术，让人文学者发怵

人文学者的数字技术短板

- 技术盲
  - 2015年12月北京大学举办的跨人文学科与数字技术的内部交流会上，北京大学信息科技学院的童云海和邓志鸿两位教授都承认自身掌握先进的数字技术，但不知道人文领域有相关需求，认为“人文学科与信息科学的研究者之间最大的问题是‘互盲’”。
- 对技术解决人文问题持怀疑和谨慎观望态度（戒心）
- 年龄分布在18~50岁之间



# 数字人文：从解决人文问题入手，让人文学者尝到甜头

- 例子：检索古文献数据库
  - 告诉他/她模糊检索

比如：

- 《论语·为政》：“子曰：攻乎异端，斯害也已矣”
  - “也已矣”是语气词，还是说“害就停止了”？

中华经典古籍库　查　“%10也已矣”
中国基本古籍库　查　“？已矣”
CTEXT 查“{1,5}也已矣”

- 其可谓至德也已矣
- 可谓好学也已矣
- 可谓明也已矣
- 可谓远也已矣
- 吾末如之何也已矣
- 亦各言其志也已矣
- 始可与言《诗》已矣

# 1、实践应该注意避免的误区三

## 误区三：自以为意做DH项目而无人文学者参与

数字技术解决人文问题，落脚点在人文

- 没有实际的人文研究和教学需求，或者需求不明确，数字人文项目实施方自以为意式的闭门造车，往往会让很多数字人文项目开始之日即是死亡之时

- 并不是所有的DH同行都能成功维持DH项目
  - 2014 Ithaka S&R Study:从1960以来在数字人文方面是公认领导者的布朗大学(Brown University)，正在逐渐失去管理层的支持。
  - 在Project Bamboo（竹子项目）———一个起初由Andrew W. Mellon基金会资助的人文cyberinfrastructure（2008年和2012年）失败以后，康奈尔大学便开始慎重的重估这项和多伦多大学共同合作的两年DH投资试点：项目的应用范围和目标不断发生变化、并没有与学者建立联系，没有建立一个共同的愿景，无法确保下一阶段的资金。

# 数字人文：让人文学者参与其中

- 不要把DH项目做成“猜谜游戏（guessing game）”
  - 图书馆的一贯风格，东西在这里，你来使用……
- 技术永远都是次要的，解决人文问题才是主要的

# 2、图书馆实践的方向

## 方向一：数字人文专题信息服务

- 对已有的DH项目/工具进行分类汇总、推介，启发本校的人文学者（可视作图书馆“学科服务”、“研究支持”的一种补充）
  - 美国北卡罗来纳大学（University of North Carolina）将本校和校外有用的数字人文项目按照人类学（Anthropology）、考古学（Archaeology）、艺术学（Arts）、古典文学（Classical Studies）、城市地理学（Geography & Urban Studies）、历史学（History）、语言文学（Literature & Languages）、哲学宗教（Philosophy & Religion）等进行分类展示，为本校的人文研究和教学提供向导；将数字人文工具按照数字人文工具包（DH Toolkits）、数据建构/清洗（Data Building/Cleaning）、线框图制作工具（Wireframing Tools）、协作工具（Collaboration Tools）、内容管理系统和网页发布工具（Content Management Systems and Web Publishing）、数据可视化工具（Data Visualization）、时间线工具（Timeline Tools）、地图工具（Mapping Tools）和社会网络分析工具（Network Analysis）等分类推介。
  - 纽约大学图书馆（New York University Library）专门将全球与英美文学（English and American Literature）相关的数字馆藏和数字人文项目专题推介出来，供全球研究者了解和使用

# 2、图书馆实践的方向

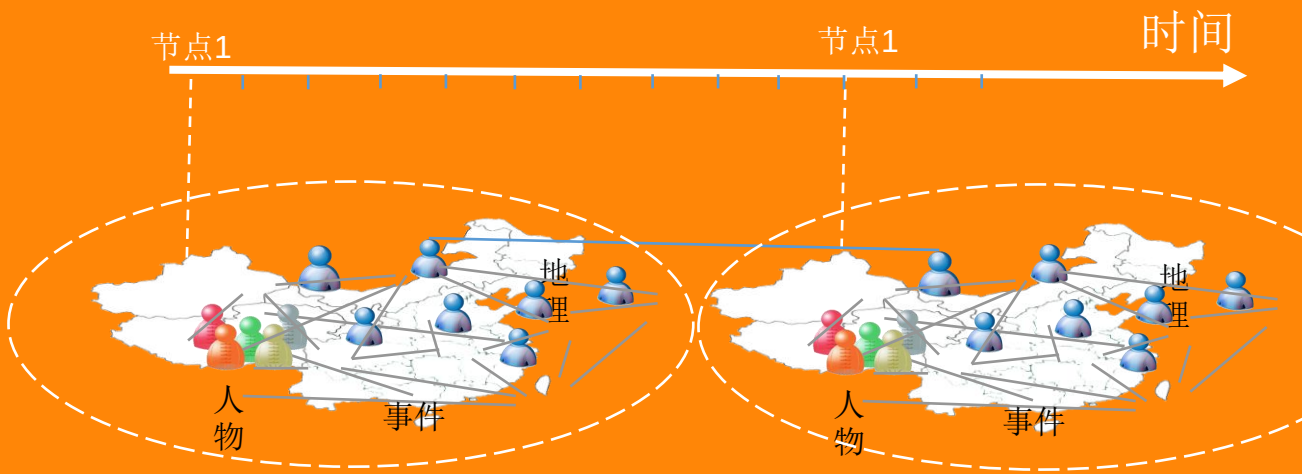
## 方向二：数字人文网络基础设施（cyberinfrastructure）

- 不只是关注于一个个资源“点”，而是一个与人文研究紧密结合且体系化的基础设施“面”
- 不只是关注于资源“卷册”、“页”信息的标引揭示，而是深入到人文内容信息的标引揭示
- 体系化地去做支持人文的，并不多见，大有空间可做

- 例子：
  - 小的项目
    - 汉典（http://www.zdict.net）、异体字字典（http://dict.variants.moe.edu.tw）、书法字典（http://www.shufazidian.com）等旨在解决汉字、汉语词汇释义、字形等查找的工具软件
  - 大的项目
    - “谷歌图书计划（http://books.google.com）”、“中国哲学书电子化计划”（http://ctext.org）、“海西数字图书馆（http://www.hathitrust.org）”、“中国历史地理信息系统（https://www.fas.harvard.edu/~chgis）”、“中国历代人物传记资料库（http://projects.iq.harvard.edu/cbdb/home）”等

# 中国史cyberinfrastructure构想

基于宏大背景的史学研究




建立时间、空间、人物、事件四个维度的cyberinfrastructure基础架构，外围做专题库

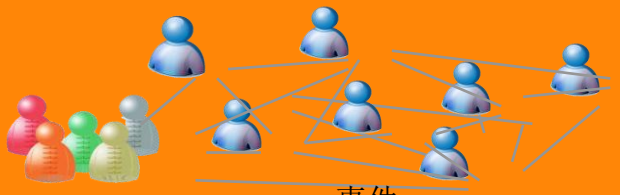
时间

- 年
  - 皇帝年号
  - 诸侯纪年（含：立、元年、二年、三年……改元、崩/薨/卒）
  - 干支年
  - ……
- 季
  - 春夏秋冬
- 月
  - 干支月
  - 十二月
  - 冬月、正月、腊月……
  - 闰（后x月）
- 日
  - 干支日
  - 朔日
  - 望日
  - ……
- 时
  - **地支时**（子丑寅卯……）
  - **五时辰**（晨明、朏明、旦明、蚤（早）食、宴（晚）食。（参阅《淮南子.天文训》）
  - 先秦**十时辰**（昼夜各五分，据《隋书.天文志》，昼为朝、禺、中、晡、夕，夜为甲、乙、丙、丁、戊（后用五更来表示）
  - 西汉**十二时辰**（夜半、鸡鸣、平旦、日出、食时、隅中、日中、日昳、晡时、日入、黄昏、人定）
  - 宋**二十四时辰**（宋以后把二十时辰中每个时辰平分为初、正两部分，这样，子初、子正、丑初、丑正……依次下去，恰为二十四时辰，同现在一天二十四小时时间一致。）
  - **五更**（一更也等于现在的二个小时，从晚上七时开始起更，一更指七时至九时，二更指九时至十一时，三更指十一时至次日凌晨一时，四更指一时至三时，五更指三时至五时）。
- 点
  - 一夜分为五更，按更击鼓报时，又把每更分为五点。每更就是一个时辰，相当于现在的两个小时，即120分钟，所以每更里的每点只占24分钟。由此可见“四更造饭，五更开船”相当于现在的“后半夜1时至3时做饭，3时至5时开船”。“五更三点”相当于现在的早晨5时又72分钟，即6时12分，“三更四点”相当于现在的午夜1时又96分钟，即2时36分。

涉及历法，古代记时与公元年月日时的换算非常复杂

# 地理

- 名称
- 沿革
  - 时间1
    - 别名
    - 地名分类
    - 行政层级
    - 首府（国；郡/道/路/州；县；乡镇）
    - 上级
    - 下辖
    - 州境
    - 四至八到（地里）
    - 物产
    - 贡赋
    - 人物
    - 户
    - 风俗
    - 舆图
    - 今经纬度
  - 时间2
    - 别名
    - ……
    - ……



事件

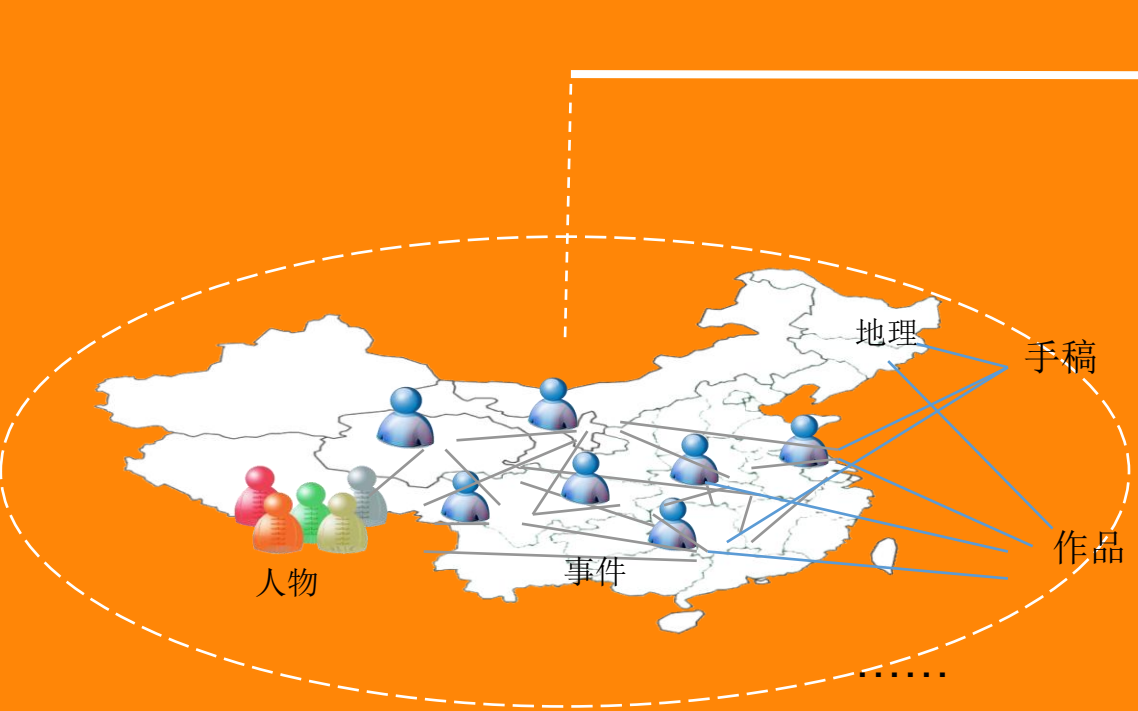
## 人物

- 人名（姓、名、字、号；谥号等）
- 时间（生卒）
- 地址
- 职官
- 入士途径（世袭；科举；察举）
- 著作
- 社会区分
- 亲属关系（自己的；配偶的；血缘关系）
- 社会关系（师生、同学、朋友、同事、同年科举）
- 事件
- 史料编年
- ……

## 事件

- 自然事件
  - 天候（日月星辰）
  - 气候（风云雷电雨雪霜露）
  - 物候（水火地动物人）
  - ……
- 人类事件
  - 杀伐
  - 交往
  - ……
  - ……

# 中国史cyberinfrastructure构想




基于SOA；Open API；

# 2、图书馆实践的方向

## 方向三：充当跨学科的桥梁

- 人文学者的数字技术短板，这个短板不仅体现在人文学者不会使用技术，而且体现在数字化方式思维人文问题，也就是可否用数字工具解决传统问题、提出新问题、解决新问题
- 图书馆可以充当“桥接者”角色
  - 开展数字人文工作坊或培训课程
  - 向人文学者普及人文领域常用的数据库、数据集、数字工具等，向拥有数字技术的工程师普及人文知识和需求，成为人文学者、计算机或信息科学人员之间的桥，翻译、整合两者之间的对话
  - 美国的西北大学（Northwestern University）、英国的牛津大学（University of Oxford）等高校，每年都会面向全校师生和校外人员提供“数字人文暑期课程”，传授相关知识技能

# 2、图书馆实践的方向

## 方向四：数字人文项目孵化器

- 在前述“桥梁”的基础上，图书馆可以提供一些“需求交流平台”进一步整合人文学者和数字技术工程师的需求，孵化新的数字人文项目或工具
- 欧美不少高校都提供了相应的孵化器平台
  - 英国伦敦大学学院（University College London）提供一个平台，既推介该校已经或正在开展的数字人文项目，又提供孵化新数字人文项目的机会
  - 美国DHCommons平台虽然不由图书馆发起，但可以为图书馆数字人文项目孵化提供一个好的样板，它为全球人文学者、数字技术工程师提供数字人文合作供给需求信息

# 北京大学数字人文做法

## “数字人文”国际学术交流平台

- 2016年5月，“跨界与融合：
- 2017年5月，“互动与共生：

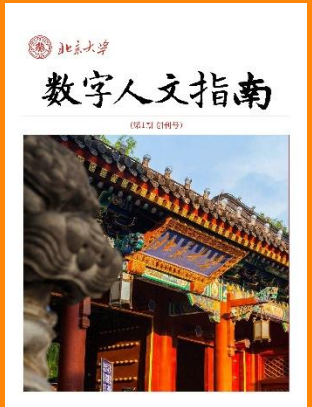
## 邀请数字人文学者做专场学术讲座

- 爱尔兰国立梅努斯大学数字人
  数字人文：新方法、新机遇和
- 台湾清华大学黄一农教授：“
- 美国伊利诺伊大学香槟分校J.
  闭和开放数据的数字人文：来

## 数字人文工作坊

- 2017年4月，第1期“历史人物
- 2017年9月，第2期“GIS与空间

## 数字人文指南




# 数字人文指南

（第1期 创刊）

主　　编：朱　强　聂　华
执行主编：朱本军
编　　委：崔海媛　罗鹏程
　　王昊贤　朱　玲
　　邹新明　秦伟平
　　钟　迪　徐蕾白
　　张　宁　孙　超
　　张乃帅　刘　丹
顾　　问：徐力恒
联系投稿：bjzhu@pku.edu.cn
　　010-62753503
印　　刷：北京大学数字加工中心

目录

# THANK YOU

朱本军　　bjzhu@pku.edu.cn