



学术图书馆语境下 数字人文的基本共识与起步工作

中山大学资讯管理学院
肖鹏（博士生）



美好愿景V.S.真实图像

- 在所有的学术探索中，“数字人文”最为生动地重现了古代图书馆的精华，它追求的是形式多样的呈现方式、知识的组织、技术的交流与传播，并致力于创造学术研究的可用工具，正是这些追求让亚历山大图书馆、帕加马图书馆、孟菲斯图书馆、中世纪修道院图书馆，乃至德国启蒙运动时期的第一批研究型图书馆生机勃勃。此外，数字人文本身就是一项学术活动——它不仅仅展现知识、发现知识，同时也关心创造知识。

——Stephen Ramsay



美好愿景V.S.真实图像

- 在图书馆中实施数字人文的挑战
- 缺乏训练的机会
- 不符合当前图书馆员职业的诉求，缺少启动方面的支持
- 过多的任务，太少的时间
- 缺少整合资源的实权
- 缺乏动机
- 与教研团队合作的复杂性
- 过分的谨慎
- 缺少的机构承诺与长期投入

——Miriam Posner



美好愿景V.S.真实图像

- 即便在起步较早的欧美学术型图书馆中，“数字人文”很大程度上还是依靠图书馆员的个人意愿和研究兴趣来推动的，仍缺乏足够的机构支持和制度保障。

图书馆员以个人研究者的身份来介入数字人文的研究：

- 基本共识：对于数字人文，我们需要知道什么，做什么？
- 起步工作：怎么做？



■ 基本共识——三观

■ 起步工作——方法论



研究方法

- 档案学大三学生（8名）的数字人文实践演练
 - 历史学者与图书馆员的访谈
 - 相关文献分析，尤其关注访谈类文本
 - 数字人文课程分析
 - 数字人文工具与工具列表分析
- 仅为初步结果而非最终结论



基本共识

- 出发点：学术兴趣反哺服务职责
- 实施中：过程性与开放性
- 段落后：保持产出、持续交流



基本共识

- 出发点：学术兴趣反哺服务职责
- “图书馆中的数字人文并非一个服务项目”的讨论（Trevor Munoz等）



基本共识

■ 出发点：学术兴趣反哺服务职责

■ 实施中：过程性与开放性

- 做完一个项目，比做一个项目更重要
- 数据处理与经验积累往往比技术体系更为重要
- 采用的技术配置一般都相当保守

基本共识

- 出发点：学术兴趣反哺服务职责
- 实施中：过程性与开放性
- 段落后：保持产出、持续交流
 - 小而美与小而丑的项目都是可以接受的
 - 交流：与上下级、与院系、与同行



补课

- 出发点：学术兴趣反哺服务职责
——馆员科研能力的培养、学科馆员
- 实施中：过程性、开放性——“开放”理念
- 段落后：保持产出——图书馆价值的证明
持续交流——学科馆员制度



起步工作

- 问题意识的明确
- 数据源的对接与创造
- 工具の利用



起步工作

- 明确的问题意识
- 模型与理论借用尤为重要



Ted Underwood的数字人文课程

- 8月23日：纵观数字人文
- 8月30日：针对数字文本（digital text）的相关工作
- 9月06日：遥读（distant reading，福朗格·莫雷蒂提出的文学理论）
- 9月13日：“模型”在人文研究中的作用
- 9月20日：计算机科学与社会理论
- 9月27日：当“遥读”偶遇机器学习
- 10月04日：数字时代的历史
- 10月11日：数字人文中的图书馆与信息科学/图书馆中的数字人文
- 10月18日：数字教学法
- 10月25日：将数字媒体化作历史资料
- 11月01日：图像与空间展示
- 11月08日：数字人文的未来
- 11月15日：解决无人教导的问题：元技能/检索“出错信息”
- 11月29日：课程展示与合作讨论
- 12月06日：课程展示



起步工作

- 研究问题的明确
- 数据源的考察与借用
- 底线是开放就好，形式并不重要

■ 利用CBDB数据库的小工具

■ 功用：让不了解数据库的人也可以定位历史人物系统编号

DiHu数字人文系统 - [基础实体]

文件(F) 编辑(E) 窗口(W) 帮助(H)

前近代中国人物 中国行政地理单位 社会关系 入仕方式 重大事件 亲属类别

姓名: 李白

Tts Sysno	C Personid	C Name	C Name Chn	Index Y	C Female
→ 26795	32540	Li Bai	李白	760	0
52744	58841	Li Baiyu	李白玉	1844	0
187813	196395	Li Bai	李白		0
	357988	Li Bai chun	李白春	1646	0

亲属关系

姓名: 李白

姓名	关系	亲戚
→ 李白	九世祖: 太高曾祖	李昂
李白	妻之祖父	許圜師
李白	族父	李陽冰

- 数据全部源于CBDB单机版，界面由肖鹏，王海明设计、开发

起步工作

■ 研究问题的明确

■ 数据源的对接与创造

■ 工具の利用

- 如何选择合用的工具：不是进行人文研究的工具，是实施项目的工具



相当宽泛的“工具”概念

工具	功用
Netvizz	从Facebook中提取数据；
Image Scraper	从单个网页中抓取（scrape）图片
Open Calais	在一组网站、一个文本或一网络中发现最为相关的词汇或词组；
DH Press	基于WordPress的非技术型用户 数字人文工具包
StoryTracker	在新闻主页上追踪消息流的工具；
Wikipedia History Flow Companion	维基百科的链接抓取与编辑应用

- Alan Liu等11个数字人文tool list的一部分，肖鹏、肖奕翻译整理



目前英文互联网所提供的工具列表

- 泛化的工具概念：生命周期管理、爬虫工具
- 泛化的数字人文概念：政治学、经济学
- 很多在中文世界并不适用性
- 必须实际参与，才有可能理解数字人文学者的需求



起步工作

- 研究问题的明确
- 数据源的对接与创造
- 工具の利用
 - 如何选择合用的工具
 - 学一点编程，才能知晓技术的边界：以R或Python为基础



一些值得做的事情

- 中文数据列表与工具列表的形成
- 网页或数据库的被引统计研究
- 二次文献/参考工具书的数字化工作



结语：回到最初？

-形式多样的呈现方式、知识的组织、技术的交流与传播，并致力于创造学术研究的可用工具.....让亚历山大图书馆、帕加马图书馆、孟菲斯图书馆、中世纪修道院图书馆，乃至德国启蒙运动时期的第一批研究型图书馆生机勃勃.....

——Stephen Ramsay

- 被忽略的语境——这些古代图书馆并无现代意义上的“图书馆员”，他们拥有纯粹的学者



结语：回到最初？

- （螺旋上升式）回归教授兼任馆员（professor-librarian）、学者型馆员的可能性？
- 中国日益增长的图书馆学博士生群体，与图书馆中日益增加的博士馆员



部分參考文獻

1. Stephen Ramsay. Care of the Soul[EB/OL]. [2016-10-20].<http://stephenramsay.us/text/2010/10/08/care-of-the-soul/>.
2. Miriam Posner. What are some challenges to doing DH in the library? [EB/OL].[2016-10-20].<http://miriamposner.com/blog/what-are-some-challenges-to-doing-dh-in-the-library/>.
3. Trevor Munoz. Digital humanity in the library isn't a service[EB/OL]].[2016-10-20]. <http://trevormunoz.com/notebook/2012/08/19/doing-dh-in-the-library.html>.
4. Chris Alen Sula.Digital humanities and libraries: a conceptual model[J].journal of Library Administration,2013.53:10-26.
5. 项洁, 翁稷安.导论——关于数字人文的思考: 理论与方法//项洁编.数字人文研究的新视野: 基础与想象[M].台北: 台湾大学出版中心, 2013: 9-20.



中山大學
SUN YAT-SEN UNIVERSITY

谢谢!

SUN YAT-SEN UNIVERSITY SUN YAT-SEN UNIVERSITY