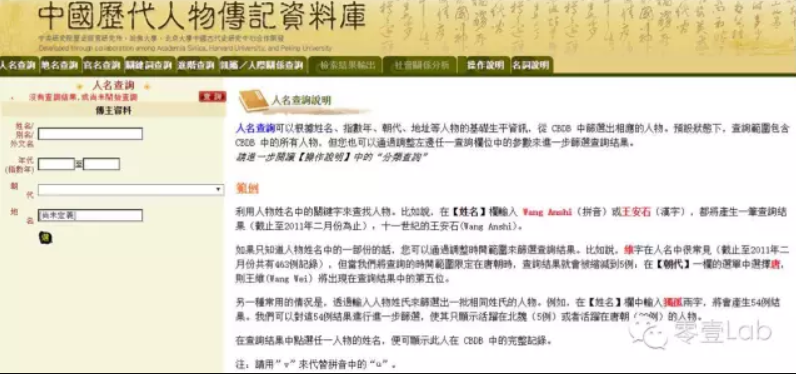
【对撞机】一个新领域的出现：中国的数字人文研究

徐力恒 [零壹Lab](javascript:void(0);) 2016-10-17



**徐力恒**  
哈佛大学博士后研究员  
“中国历代人物传记资料库”目成员  
电邮地址为：tsui01@fas.harvard.edu

“数字人文”研究（台湾一般称“数位人文”，香港有时称作“数码人文”）就全球研究动向而言不算一件全新的事物，但最近几年才开始受到中国学界的高度关注。项洁、王晓光、陈静等几位学者已分别对中国的数字人文现况做过介绍，以下我希望就自己接触到的情形，提出一己之见。中国高校的第一个数字人文中心是武汉大学在2011年建立的，至今它仍是中国大陆地区唯一的数字人文中心，是数字人文中心网络（centerNet）旗下亚太数字人文中心网络（Asia-Pacific DH centerNet）的成员。在组织形式上稍有不同的另一机构是同在中国大陆的南京大学的“数字人文与超媒体GIS工作室”，着重对历史地理信息系统的建设和研究。台湾则主要有台大长期投入数位人文的研究，2007年已成立“数位典藏研究发展中心”，2012年改名为“数位人文研究中心”；而政治大学既有数位人文计划，沟通校内的数字人文研究，也成立了文学院下属的“数位人文研究中心”。香港的公开大学也在2015年成立了“数码文化与人文学科研究所”，推动数字人文的学术交流。  
  
虽然相关的研究中心为数不多，但投入数字人文研究的项目和专家还是不少。**早在二十世纪七八十年代，中国学者对语言的研究就推动了不少现在可以认为属于数字人文的学术项目**，例如1976年武汉大学语言自动处理研究组利用计算机统计老舍《骆驼祥子》字频；1979年以来为了分析现代汉语，出现了好几种机器可读的语料库等。由于中文信息处理需要对汉语的特点和人文意涵进行研究，又必须有计算机技术上的创新作为支撑，这就理所当然地造就了数字和人文的融合。  
  
一些高校近两年举办的活动也反映人们对数字人文的关注；比如北大图书馆在2016年5月举办的首届“数字人文论坛”，得到超过150位学者和学子参加，反映出人们对这块新兴领域的浓厚兴趣。各人文领域的学者至少明确意识到，“数字人文”作为一个新概念可以引导不同学科的专家反思数字化工具带来的机遇和冲击，所以几个学科都举办了活动邀请学者进行讨论，例如是图书馆学、清史研究等等。当然，除了讨论以外，一些高校也在数字人文范畴投放资源，从事具体的研究，例如赞助历史地理信息系统（GIS）的建设。**另外，可以预见的是高校图书馆以后在数字人文研究中会扮演非常重要的角色，因为它们既有管理甚至是创建数字资源的经验，又需要面对读者、用户和研究者，可能最具备担当桥梁角色的条件。**它们和科研人员的合作会是推动数字学术（digital scholarship）发展的关键。  
  
**数字人文的研究项目往往是集体协作、国际合作的，**有些中国的大学就和海外的研究机构从事这方面的学术合作。例如，我供职的“中国历代人物传记资料库”（CBDB）就是一个已经运作超过十年的国际合作项目，由哈佛大学、中央研究院和北大共同开发。这数据库目标在于系统地收录中国历史上所有重要的传记资料，并将数据开放供学术研究之用。截至 2016年，本数据库共收录超过37万人的传记资料，这些人物主要来自7到19世纪。数据库的数据既可在线查询，又可以全部下载，供用户离线使用。研究者可以利用其中提供的大数据，进行相对复杂的查询和分析——除了用作研究历史人物的参考资料之外，还可作统计分析、地理空间分析与社会网络分析之用，为中国史研究引入新视角。从2016年起，这个数据库项目在中国连续举办了不少推广活动，向学界介绍其资料特点和用法。



（“中国历代人物传记资料库”的线上查询界面）

此外，还有为数不少的研究和电子化项目，比如许多国家社科基金项目都资助学者建设各种专题数据库。然而，不少学界同仁还觉得无从入手学习数字人文的最新动态，认为各个学术机构还可以投放更多资源，让研究者学会如何在研究中利用新的数字化工具。比如，人文学者通常都熟悉在全文数据库进行关键词检索，但对于其他可以用于研究的计算机工具，一般还是很陌生的。例如，要把自己搜集到的数据以GIS 方法画一张电子地图，就不是很多人能够做到。所以，**推动数字人文的发展，与其停留在讨论数字人文的理念，或介绍众多数据库和电子资源，还不如注重实践，例如培养制作可视化的技能，或传授如何对数据进行分析、操作、解读等技能。**例如，面对充斥媒体报道和学术报告、论文的可视化图像，我们需要带着什么意识去解读和提问？要自己制作的时候又需要注意什么？诚然，不是每一位人文学者都要系统地学习数据科学的技术和方法，但不管是否用于自己研究，他们每位都值得接触数字人文的前沿研究方法，对其进行反思。连带的一个问题是，人文社科研究生的培养如何应对数字人文带来的创新和范式，也成为学术共同体不得不面对的问题。我们不能被动地认为在数字化环境下土生土长的数字原生代（digital natives）一定具备这样的意识，其实他们也是需要培训和引导的。  
  
**数据的获取和开放程度也是中国数字人文面临的另一大挑战。**以中国古代典籍为例，电子化材料的获得远远不是开放的。各类古籍数据库多如牛毛，但数据共享的做法仍然非常罕见——许多数据库都以商业模式运营，必须得到学术机构和研究者订购，才能生存。因此，它们的数据肯定不是完全开放的，这对不同电子资源之间的协作造成一定障碍。对于费用高昂的数据库，不少学校不能负担，也是另一难题。  
  
虽然如此，还是有一些机构希望推动开放数据的做法。例如上海图书馆就建立了开放数据平台，以关联数据（linked data）的方式发布一些各个机构、项目都可调用的数据。同时，又创办了应用开发竞赛，开放了其馆藏家谱文献信息和内容信息，鼓励参加者有创意地利用数据，从而发挥资源的最大价值。类似活动无疑有利于推广数字项目，让更多人了解数字人文的理念和成果。对数字人文感兴趣的学者也越来越习惯在网络上讨论该领域的相关议题，尤其是用微信沟通，交流对各种电子资源、项目的意见，或发布相关学术活动的信息。一些微信公众号和群组也专门用于讨论数字人文研究（比如您正在看的这个“零壹Lab”！）。即便数字人文在学术研究的正统框架内还没有足够的影响力，许多交流已经通过非正式、非传统的渠道进行。



（上海图书馆的开放数据应用开发竞赛）

就如其他地区的学者一样，参与数字人文研究的中国学者感到不同项目和机构之间应该有更紧密的联系和合作。**随着项目数量和参与人员的大幅增加，为中国的数字人文建立相应的网络基础设施（cyberinfrastructure）变得很有必要。**这样的基础设施可以连接不同项目和机构，尤其是他们开发的计算机工具和数据集，也可以促进各种参与者的沟通，为建立共通的标准和方法努力。只要项目之间有足够的协调，就不用再浪费资源在重复的工作之上，例如多个机构为版本完全相同的书进行电子化。越来越普遍的应用程序接口（API）也有助项目之间的连接和灵活地进行跨项目的操作，甚至是使用一些共同的工具。不管是“中国历代人物传记资料库”还是“中国哲学书电子化计划”（CTEXT），都应用程序接口使得其他数据库和系统得以调用人物或古籍数据，以不同的方式灵活呈现。总而言之，共同建立一个连接不同项目和专家的技术设施，对从事数字人文的所有人都是有利的。我们希望这样的对话能尽快开始——当研究者开始商量怎么合作时，数字人文作为一个领域就算正式在中国出现了。  
  
（本文原以英文在诺丁汉大学中国政策研究所博客发表，本篇中文版经过改写和补充。）

推荐阅读

项洁、涂丰恩，〈导论——什么是数位人文〉，收于项洁编，《从保存到创造：开启数字人文研究》（台北：国立台湾大学出版中心，2011），页9-28，http://thdl.ntu.edu.tw/thdl\_pdf/book1.pdf

相关链接

**项洁教授简介**：http://www.digital.ntu.edu.tw/hsiang/

**王晓光教授简介**：http://sim.whu.edu.cn/sz/jsxq/4/2016-05-09/873.html

**陈静教授简介**：http://artint.nju.edu.cn/TeacherDetail.aspx?newstype=9&id=32

**武汉大学数字人文研究中心**：http://sim.whu.edu.cn/xwdt/2016-05-25/1274.html

**数字人文中心网络（centerNet）**：http://dhcenternet.org/  
**南京大学“数字人文与超媒体GIS工作室”**：http://114.212.112.28:81/index.php

**台湾大学数位人文研究中心**：http://www.digital.ntu.edu.tw/index.jsp

**政治大学数位人文计划**：https://sites.google.com/site/nccuwiki/home

**政治大学文学院“数位人文研究中心”**：http://www.la.nccu.edu.tw/allCenter1/pages.php?ID=allcenter1

**香港公开大学数码文化与人文学科研究所**：http://www.ouhk.edu.hk/wcsprd/Satellite?pagename=OUHK/tcSchWeb2014&l=C\_ASS&lid=1385176377965&c=C\_ASS&cid=1385176376066&lang=chi&sch=ASS&mid=0

**中文信息处理介绍**：http://202.106.125.14:1168/indexengine/entry\_browse.cbs?db=book1&value=%D6%D0%CE%C4%D0%C5%CF%A2%B4%A6%C0%ED&jm=%D3%EF%D1%D4%A1%A4%CE%C4%D7%D6

**首届北京大学“数字人文论坛”**：http://pkunews.pku.edu.cn/xwzh/2016-05/25/content\_293906.htm

**2014年图书馆前沿技术论坛**：http://library.gensee.com/webcast/site/vod/play-ad9404318ac447aa98bfcf0a33c5b16d

**“数字人文与清史研究”学术工作坊**：http://pkunews.pku.edu.cn/xwzh/2016-05/25/content\_293906.htm

**“中国历代人物传记资料库”（CBDB）**：http://projects.iq.harvard.edu/chinesecbdb

**中国历代人物传记资料库数据在线查询**：http://db1.ihp.sinica.edu.tw/cbdbc/ttsweb?@0:0:1:cbdbkmeng@@0.6341087994044654

**中国历代人物传记资料库数据下载**：http://projects.iq.harvard.edu/chinesecbdb/%E4%B8%8B%E8%BC%89cbdb%E5%96%AE%E6%A9%9F%E7%89%88

**中国历代人物传记资料库推广活动**：http://projects.iq.harvard.edu/chinesecbdb/%E4%B8%96%E7%95%8C%E5%90%84%E5%9C%B0%E7%9A%84%E4%B8%AD%E5%9C%8B%E6%AD%B7%E4%BB%A3%E4%BA%BA%E7%89%A9%E5%82%B3%E8%A8%98%E8%B3%87%E6%96%99%E5%BA%AB%E7%A0%94%E8%A8%8E%E6%9C%83

**上海图书馆开放数据平台**：http://data.library.sh.cn/

**开放数据应用开发竞赛**：http://pcrc.library.sh.cn/zt/opendata/

**ACLS Commission on Cyberinfrastructure**：https://www.acls.org/programs/Default.aspx?id=644

**Crowdsourcing, APIs, and a Digital Library of Chinese**：https://cpianalysis.org/2016/06/08/crowdsourcing-apis-and-a-digital-library-of-chinese/

**应用程序接口介绍**：http://baike.baidu.com/view/592964.htm

**“中国哲学书电子化计划”（CTEXT）**：http://ctext.org/zh

翻译 / 陈夏辛

责编 / 徐力恒 陈夏辛 顾佳蕙 彭玮 桑雨茜

美编 / 曹清语