

## 破除学科旧壁垒,赋予前进新动能

——“破壁与赋能:多学科驱动下的数字人文”国际学术研讨会综述

吕金伟 朱锁玲

(南京农业大学 中华农业文明研究院,江苏 南京 210095)

【中图分类号】S-09;K207 【文献标志码】A 【文章编号】1000-4459(2021)04-0140-06

数字人文研究蓬勃发展,逐渐成为显学。然而,在繁荣景象的背后,不同学科之间的壁垒依然难以打破,研究共识尚未凝聚,跨界合作仍待加强。为了探索“数字”与“人文”之间的有机融合之道,2021年5月21—23日,“破壁与赋能:多学科驱动下的数字人文”国际学术研讨会在南京农业大学学术交流中心召开。本次会议由南京农业大学主办,南京农业大学数字人文研究中心承办,西北大学科学史高等研究院、上海图书馆历史人文大数据中心、中国索引学会数字人文专业委员会、中国科学技术史学会数字人文专业委员会(筹)、上海市高等院校海外交流联谊会新文科专业委员会协办。来自北京大学、复旦大学、南京大学、南开大学、武汉大学、西北大学、北京科技大学、上海图书馆、哈佛大学、莱顿大学、杜伦大学、南京农业大学等25家国内外高校和科研机构的100余位代表参加了会议,2600余位学者通过腾讯会议在线观看了会议直播。

开幕式上,南京农业大学党委常委、副校长胡锋,江苏省哲学社会科学规划办公室主任许益军分别致辞。胡锋代表南京农业大学致欢迎辞,他指出,南京农业大学长期支持和鼓励开展交叉学科研究,数字人文是新兴交叉学科的代表,“数字+”为传统人文学科发展过程中遇到的问题提供了新的解决方案,“人文+”为数字技术在人文学科中的应用提供了广阔天地。“数字”与“人文”的有机融合,既能使传统人文研究的厚重感更加鲜活,又能使现代数字技术的机械感更具温度。许益军在致辞中强调,南京农业大学召开本次会议是见证数字人文研究蓬勃发展的一件盛事,充分体现了专家学者对数字中国战略的深入理解,对科研需求的准确把握,对学术潮流的努力追赶。数字人文研究借助“数字”技术解决“人文”问题,是人文研究在新时代的新发展,对推动中国特色哲学社会科学创新发展具有十分重要的意义。

围绕“破壁与赋能:多学科驱动下的数字人文”这一主题,与会专家学者从历史学、科技史、图书情报学、古典文献学、艺术学、计算机科学等学科角度展开了广泛而热烈的探讨。会议以专家报告与圆桌论坛相结合的形式进行,根据大会主题,专家报告主要分为四个方面:面向(史学)领域的数字人文研究需求与困境;面向(史学)领域的数字人文研究基础设施建设;数字人文研究方法与应用、效果与局限;数字人文背景下多学科的融合与边界。圆桌论坛以数字技术与人文研究如何“破壁”、如何“赋能”为切入点,与会专家学者进行了面对面的深入交流。

【收稿日期】2021-06-17

【基金项目】国家社会科学基金重大项目“方志物产知识库构建及深度利用研究”(18ZDA327)

【作者简介】吕金伟(1990—),男,南京农业大学中华农业文明研究院博士后,研究方向为科技史;

朱锁玲(1983—),女,南京农业大学中华农业文明研究院副教授,研究方向为古籍数字化、农业文化遗产保护。

## 一、面向(史学)领域的数字人文研究需求与困境

数字人文发端于20世纪40年代末以来的人文计算,是一个由人文知识、计算机网络基础设施、数据分析与可视化技术等多种技术和知识融合发展形成的新的跨学科研究领域,是数字时代新型知识生产范式的代表。2019年,“大数据视域下数字人文研究”入选“2018年度中国十大学术热点”。大数据时代的到来,广泛而深刻地影响着以史学为代表的传统人文研究的发展,数字人文逐渐受到史学研究者的关注。

### (一)数字人文与农史研究

自1920年金陵大学农业图书合作部的筹建开始,规模化的农史研究逐渐在中国兴起,至今已有一百余年的历史。其中,农史研究资料的搜集与整理作为中国农史研究的主要方向之一,大致经历了纸本化、电子化、数据化三个阶段。大数据时代背景下,数字人文的迅速发展给农史研究带来了新的机遇与挑战,数字人文背景下的农史研究逐渐兴起,它是指借助数字人文的理论与方法,探讨农史研究中的相关问题,实现数字人文提倡的“数据驱动”与农史研究标榜的“问题驱动”的有机结合。

南京农业大学包平教授及其团队近年来一直致力于以地方志物产为核心的数字人文基础设施建设与深度利用研究。包平教授在《方志物产数字人文研究实践与思考》的报告中指出,在万国鼎先生组织辑录《方志物产》手抄本、王思明教授主持进行《方志物产》影像化与文本电子化、侯汉清教授引导开展农业古籍智能整理等工作的基础上,方志物产的数字人文研究得以持续深入地推进;在小规模语料测试成功的基础上,基于机器学习、深度学习等方法开展大规模方志物产文本(约3600万字)的自动分词与词性标注、自动断句与标点、命名实体识别等研究,借助社会网络分析软件研究物产与人物、时间、地点、医疗功效之间的关系,构建基于关联数据四原则与语义技术框架的方志物产知识库;尝试开辟和挖掘地方志物产知识在地理标志产品的历史文化内涵、传统种质资源的普查与保护、药用物产的智能整理与利用等方面的应用场景。同时,他也认识到地方志物产数字人文研究中仍存在资料是否完整、集外字如何处理、同物异名与同名异物现象如何通过智能方法进行考证、文献视角的物产变迁与基因视角的生物演进之间如何印证等难题,进一步的资料再辑、面向应用的方法体系构建等依然需要持续深入的研究。

### (二)数字人文与科技史研究

科技史是一门关于科学技术产生、发展及其变化规律的交叉学科,数字人文的兴起为科技史研究提供了新的研究范式。北京科技大学潜伟教授在《科技史研究的数字化趋势》报告中谈到科技史研究中存在数字化缺位的现象,但计量史学、数字史学、e考据等史学思潮也在一定程度上影响着科技史研究中数字化趋势的发展。他认为科技史研究的数字化大致会经历三个阶段,即科技史史料数字化存储或检索管理、科技史史料的可视化展示、基于大规模历史数据挖掘的量化分析研究,从单纯时间坐标轴到时空多维分布、从人工统计到计算机辅助信息管理、从简单数值计算到大数据人工智能、从大科学史研究到精确科技史研究是未来科技史研究中数字化的发展趋势。

西北大学曲安京教授做了题为《中国出土文献的数字人文研究》的报告,介绍了“中国出土文献语料库”的构建与应用,指出“中国出土文献语料库”囊括甲骨文、金文、简帛等出土文献,已经在分词标注、词性标注、集外字、数据可视化、出土文献的复原与缀连、短文本的处理上取得一定的成效。他以里耶秦简为例,谈到数字人文研究方法的应用首先要结合里耶秦简自身的特点,利用自然语言处理技术对文本进行预处理;然后通过制定文本录入规范和分词规范,利用和创新现有的分词算法,找到适用于里耶秦简的解决方案。最后,他强调将数字人文研究方法应用到出土文献的研究领域,可以为出土文献的研究提供新方法,在中文信息处理与历史问题诠释上发挥现代信息技术的优势,推动包括农史、数学史在内的

科技史研究。

### (三)数字人文与历史研究

数字技术是一把双刃剑,它在推动史学研究发展的同时,也给史学研究带来了一些问题。哈佛大学包弼德教授在《数字环境下的研究周期:面临何种挑战?》报告中阐述了研究周期的内涵,认为研究周期包括查找资料(或数据、信息)、整理资料、分析资料、出版资料、研究问题等环节;在查找资料方面,Google Book Search、ARTstor、中国哲学书电子化计划、爱如生典籍数字平台等数据库的搭建为此提供了便利;在整理资料方面,数字技术改变了资料存储、共享的模式,但同样存在着存储空间有限、数据共享意愿较低、服务器断网等问题;在分析资料方面,文本挖掘/文本建模、基于关系型资料的建模、群体传记学、空间分析、社会网络分析等概念的演进,使基于CBDB的分析呈现出不同的面向;在出版资料方面,开放获取期刊和文章变得越来越便捷。

数字人文是否会瓦解历史学甚至淘汰历史学家?这是一个饶富意味的话题。南开大学王利华教授在《数字资源与数字陷阱:基于研究实践的反思》报告中分析到,数字技术一方面改变了史料搜集、存储、检索、利用的方式,有助于史学研究的展开。他以20世纪90年代自己从事《广志》的辑佚工作为例,设想如果放在今天,数字资源的获取将大大提高辑佚工作的效率。另一方面,他也清醒地认识到数字技术不可能完全替代人脑的思考,对经过检索得来的史料进行脱离语境的统计、分析,容易出现断章取义的问题。他以《全唐诗》中松、竹、茶、鱼、鸟、羊、猪、狗等物产为例,阐述了同物异名与同名异物的考证需要回到具体的历史语境中进行合理解读的观点。

## 二、面向(史学)领域的数字人文研究基础设施建设

基础设施建设是数字人文研究得以展开的前提条件,包括数据库、知识库、操作平台、可视化平台、众包应用平台等,一般由图情、计算机领域学者组织、设计与开发,为人文学者获取资料、分析资料、利用资料提供帮助。

云时代的到来,为各种基础设施之间实现资源共建共享带来了新的机遇与挑战。上海图书馆刘炜研究员在《数字人文平台架构及语义互操作设计》报告中指出,目前数字人文的发展处在瓶颈期,数字人文基础设施建设已经卓有成效,但每一个基础设施依然是单独的系统,尚未实现互联;进入云时代,由系统(技术)架构与内容(语义)架构共同组成的数字人文平台可以通过API接口在互联网上进行互操作,其中,系统架构的应用场景包括知识单元的标识及其管理、支持多种协议的跨网域搜索发现或获取链接、微服务的容器及编排规范、多种数据类型的管理、复杂但统一的用户及授权管理等,内容架构的应用场景包括一致性/相似性计算、工作流定义对研究流程的支持、各类图像功能的支持、文本与图像关联、提供证据链服务、海量数据可视化支持、事实的可信度计算及排序、众包数据加工平台的数据管理、数据系统迭代进化的支持等。

文化遗产分为物质文化遗产与非物质文化遗产,对其进行数字化是当前数字人文领域的研究热点之一。武汉大学王晓光教授在《文化遗产智能计算:方向与路径》报告中介绍了文化遗产智能计算的发展背景、概念、意义、典型案例、研究方向、研究目标等内容,认为文化遗产智能计算是指借助大数据、人工智能、云计算、5G等前沿数字技术对文化遗产蕴含的信息与知识进行采集、分析、组织、挖掘、表达、传播、展示,促进物质和非物质文化遗产从现实世界向数字空间迁移,实现文化遗产的数字化再造与活化,探索文化遗产大数据驱动的数字人文研究,以及人类文明的数字化转型。其研究方向包括文化遗产智慧数据建设与应用、古籍智能化整理、挖掘与再造、文化遗产图像语义理解与计算、文化遗产GIS、文化遗产3D建模与虚拟呈现等五个方面。他还强调,文化遗产智能计算需要多学科共同参与,是开启数字



文化新时空和数字文明的关键。

图像是史料的重要组成部分,但其价值长期以来被低估,当前图像史学的兴起有助于这一现象的改变。上海图书馆陈涛博士在《史料资源图像知识框架构建与应用——MISS,从图像到智慧》报告中阐述了数字人文在图像知识挖掘、分析中的应用,认为目前图像更多的是作为“附件”而非“资源”存在,通过国际图像互操作框架(IIIF),可以实现图像的互操作、可获取、可展示、可关联、可比较、可分析,使图像成为可以流通、共享、交互的数据节点。他还介绍到,依据国际图像互操作框架(IIIF)等标准构建的多维度图像智慧系统(MISS平台)支持多种格式的图像资源一站式在线组织、发布、复用、语义标注、分享等,实现了图像数据库结构存储、遥读/近读、内容重组、OCR、多层标注、协同研究、对象识别/关联等功能,为耕织图、山水画、人物画等多种图像资源的利用提供了坚实的技术支撑。

### 三、数字人文研究方法与应用、效果与局限

数字人文的研究方法多样,如文本分析、社会网络分析、GIS可视化等,针对不同的学术问题采取不同的研究方法,才能合理的解决问题。然而,值得注意的是,数字人文研究方法的应用也无法解决所有问题,或多或少带有一定的局限。

数字人文在史学研究中的应用能够起到什么样的效果呢?北京大学王军教授在《数字人文能为历史研究做什么?》报告中强调,数字人文可以为史学工作者提供一些辅助工具,但无法替代史学工作者的思考。在大数据时代,数据库的核心功能在于提供大规模的存储与快速搜索,史料的极大丰富也带来了如何阅读的难题,为了解决这一问题,文本材料的数据化首先要进行专有名词(时间、地点、人物、事件、主题等)的识别;同时,人工智能的发展,使计算机处理史料的能力从搜索迈向理解。据他介绍,北京大学设计开发的吾与点古籍自动整理系统对包括古农书在内的古籍的自动断句、命名实体识别较之人工具有一定的优势;宋元学案知识图谱通过人物、时间、地点、著作、职官等信息的展示,为观察宋元时期的学术发展提供了动态景象。

人口问题是关系中国历史的大问题,学界讨论热烈,聚讼不已。数字人文研究方法的应用,有望为一些争论的解决提供可能。复旦大学路伟东教授在《数字人文背景下的长时段中国历史人口研究》报告中以中国传统时代后期的城市人口规模、城市化水平、静态数据展示的人口动态迁移、《中国人口史》五六卷数据检验等问题为分析对象,对以往学者的观点进行评议,并以宣统“地理调查表”为据,基于随机森林回归模型等方法,重建了清代西北地区人口空间分布格局。他表示,数字人文是大势所趋,为了避免重走计量史学的老路,数字人文背景下的史学研究需要高扬人文的光辉,切忌用单一的“数据变量”取代构成历史主体的人的主动“历史思考”和研究者本身对历史复杂性的个性化全面理解和整体认知。

数字人文研究方法的应用,一方面确实推动了包括史学研究在内人文研究范式的革新,另一方面也不可避免地遇到了一些阻碍。莱顿大学魏希德教授在《东亚数字人文的关键问题》报告中归纳了近年来东亚数字人文发展中出现的五种趋势:其一,产生了一大批代表性成果,如CBDB、CHGIS、C-Text等资料库,MARKUS文本标记平台、COMPARATIVUS文本对比平台、PARALLELS版本对比平台等研究平台,RISE等不同机构研究者共同设计开发的资源搜索平台,以及数字人文发展速度越来越快、研究呈现多样化、虚拟学术共同体逐渐形成等;其二,存在着来自于学术界内部的障碍(壁垒)、技术性的挑战、资料管制的挑战、认识论的挑战等难题;其三,数字人文的不同理念、不同制度嵌入对国际合作存在着影响,其中,面向计算机领域的研究与面向理论的研究并不均衡,后者较为少见;其四,国内、国际的标准不同带来了一定的问题,未来的基础设施建设与研究需要在共同的标准下开展工作;其五,数字人文的教学工作较之研究工作相对薄弱,教材、课程设置等问题亟待解决。

#### 四、数字人文背景下多学科的融合与边界

数字人文属于跨学科的交叉研究,需要多学科的协同推进与共同参与。然而,来自于不同学科、具有不同知识背景的学者们往往难以实现数字技术与人文素养之间的完全贯通,以致常常出现数据驱动与问题驱动两种截然相反的研究取向。因此,不同学科之间如何进行融合、融合程度如何、是否存在壁垒、如何破除边界,都是数字人文领域值得探讨的话题。

南京大学陈静副教授在《迭代更新还是打破重构? 数字人文之于交叉学科的价值》报告中谈到,交叉学科的研究正在日益兴起,以往的经验显示,它并无单一的研究模式,开展起来也存在着不小的难度。在理工科领域,生物医学、物理化学、天文物理以及统计学与其他学科之间的交叉得到了较好的发展;在文科领域,文化研究、性别研究、审美人类学、艺术社会学等积累了一定的成果。其中,文化研究这一研究领域的兴衰为当前数字人文的发展提供了可资借鉴的参考。她还认为,数字人文在应对新一轮数字技术变革带来的挑战、响应国家主导的新文科战略、实现学科建设的内部突破、构建新的学术共同体与评价体系等方面具有重大意义。

杜伦大学德龙助理教授做了《从计算机科学和中国历史研究的角度反思数字与人文的相遇》的报告,他从自己的教学经历出发,对比分析了研修《中国研究的数字方法》课程的具有人文背景的硕士研究生与研修《人文、社会科学中的计算模型》课程的计算机系本科生之间的区别,认为前者具有十分强烈的问题意识,但会遇到缺乏计算机思维、难以评估方法应用的难度等概念上的挑战,以及需要从零基础开始学习编程、如何高效地利用相关工具等实用上的挑战,后者则善于从数据或者工具出发思考问题,但也会遇到受限于计算机思维、不善于评估何种分析值得做等概念上的挑战,以及难以找到适当的研究空白、经常选择保守的选题等实用上的挑战。他还期望,“数字”与“人文”之间可以产生更进一步的交流,进而推动数字人文的继续发展。

#### 五、圆桌论坛:如何“破壁”与“赋能”?

数字人文力图弥合“数字”与“人文”之间的鸿沟,架起“数字”与“人文”之间的桥梁,实现“数字”与“人文”之间的融合。所谓“破壁”,是指破除“数字”与“人文”之间的学科壁垒。所谓“赋能”,是指通过“数字”与“人文”之间的相互借鉴,赋予彼此新的前进动能。唯有破除“数字”与“人文”之间跨界研究的旧障碍,才能建立“数字”与“人文”之间有机结合的新局面。

##### (一)人文学者的思考

围绕着数字人文学科化的潜力与阻力、数字化过度强调量化研究是否会使史学研究失去温度等问题,与会人文学者各抒高见。

南开大学王利华教授强调,数字化时代的到来是不可逆转的趋势,数字技术的迅速发展是否会造成人文学者在追赶这股时代潮流的过程中迷失方向是一个值得关注的问题,只有实现工具理性与人文情怀之间的有机结合,才能真正推动数字人文背景下史学研究(农业史、环境史等)范式的变革。

复旦大学路伟东教授谈到,正是由于图书馆、博物馆、档案馆工作人员的辛勤努力,今天的人文学者才能够便捷地利用越来越多的网络资源,并在此基础上展开研究。他认为,人文学者对资料的利用经历了能用、精用两个阶段,借助数字技术处理数据并不一定会消解史学研究的人文情怀。

北京科技大学潜伟教授指出,数字人文作为一个新的交叉研究领域,从学科建制上来看,尚不具备成为新学科的条件。他还讲到,构建资料丰富、可信的数据库是数字人文研究的第一步,利用数字技术

进行分析可以为科技史研究提供帮助,而非解决所有问题。

西北大学曲安京教授认为,数字人文是当前学术研究的热门领域,但却存在着兴勃亡忽的危险,如果仅仅满足于将学术问题通过复杂的、生动的图景展示出来,数字人文的意义会大为削减,能否解决问题才是数字人文未来存在与否的唯一标准;基于解决问题的角度而言,数字人文只是一种辅助手段,数字人文背景下的史学研究不会颠覆传统史学。

## (二)图情学者的思考

围绕着数字人文研究是否存在中心化的趋势、数字人文研究者应当具备何种素养等问题,与会图情学者畅谈宏论。

武汉大学王晓光教授指出,当前对数字人文的认知存在着工具论、主体论两种倾向,前者视数字技术(资源)为支撑人文研究的一种工具,后者则关注数字技术带来的社会问题,唯有技术与人文之间形成良性的互动才能推动数字人文的发展。他还表示,培养兼备数字技术与人文知识的跨学科人才是未来数字人文领域人才培养的重要目标。

上海图书馆刘炜研究员强调,数字人文的发展是大势所趋,跨学科的交流与融合是十分必要的,由应用数字人文、理论数字人文与技术数字人文组成的数字人文仍然需要来自不同学科的研究者们共同推动。他满怀信心地期待,通过图情学者与人文学者的亲密合作,数字人文的内核正在逐渐形成,学科化的建制也会不断推进。

北京大学王军教授认为,面对数字人文蓬勃发展的时代潮流,来自数字技术与人文研究两个阵营的学者都需要做出一定的改变,其中,人文学者应当抱持接纳数字化趋势的态度,关注新问题,而非过分苛责数字人文无法解决传统人文研究遗留下来的难题;数字技术专家要不断培养人文素养,对人文研究怀有敬畏之心,时刻牢记技术并不是万能的。

南京农业大学包平教授总结道:首先,当前数字人文领域缺少争鸣与批判,恰是蓬勃发展背景下需要冷静思考的一个问题;其次,数字技术与人文研究需要开展更多、更为广泛地合作,任何领域的学者都不应该固步自封;最后,数字技术的发展日新月异,能够解决问题的方法越来越多,能够解决的问题也越来越多,人文学者积极地拥抱技术进步,将有助于他们更好地解决问题。

伴随着信息技术的不断发展,数字化正在深刻影响着当今社会的方方面面。数字人文的兴起,正是学术界对这一时代潮流的回应与追赶。在人文、图情、计算机等相关领域学者的共同努力下,数字人文已在基础设施建设、大数据分析、长时段的宏观历史解释、GIS可视化呈现等方面取得突出成效。针对其间存在的问题,相关领域学者也已有清醒的认识,并在积极地寻求解决方案。以本次研讨会的召开为契机,数字人文背景下的农史研究将在充分考量人文学者需求的基础上,通过发挥数字技术的优势,继续推进农史文献智能整理、农史数据库(知识库)建设、农业文化遗产(农史文献、耕织图、农业生产技术、农业工具与器械、水利工程、农业遗址、农业聚落、农业景观等)数字化、农史知识聚合、农史知识发现、农史知识应用等工作,开拓农史研究的新领域,引入农史研究的新方法(新工具),解决农史研究的新问题。数字人文的深入发展正在推动农史研究的方法创新,将为农史事业的发展壮大而贡献力量。

(责任编辑:徐定懿)