日本数字人文项目探析和启示*

徐彤阳　顾婷婷

摘　要　运用网络调查法和邮件访谈法梳理日本数字人文项目整体概况,按研究主题对研究项目进行学科和内容分类并介绍其代表性机构及项目成果,从中总结日本数字人文项目在研究主题、研究深度和广度、数据开放化程度以及研究重点等方面的特征,最后从对数字人文研究的支持力度、学科融合交流合作、研究方向、研究层次四个方面对我国数字人文发展提出参考意见。图1。表3。参考文献68。
关键词　日本　数字人文项目　实践

Analysis and Enlightenment of the Characteristics of Digital Humanities Project in Japan

Xu Tongyang　Gu Tingting

Abstract: This paper uses the method of network survey and interview via email to introduce the overview of the Japanese digital humanities project, classifies the projects according to the research subject and content of research project, introduces the representative institutions and projects, summarizes the characteristics of Japanese digital humanities projects from research topic, research project in depth and breadth, data open degree and the key aspects of study, thus to finally provide referential suggestions for the development of digital humanities research in China from four aspects: strengthening the support of digital humanities research, focusing on subject fusion exchanges and cooperation, broadening research scope, and deepening research level. 1 fig. 3 tabs. 68 refs.
Keywords: Japan; Digital Humanities Project; Practice

数字人文(Digital Humanities)是将计算机技术与传统人文学研究相结合的新型跨学科研究领域,其概念由“人文计算”(Humanities Computing)发展而来[1]。早期的人文计算主要应用在文学和语言学领域,随着计算机技术的发展,人文计算渗透到历史、文化、艺术等多个领域。2004年,在《数字人文指南》(*A Companion to Digital Humanities*)一书中首次出现了“数字人文”概念[2]。数字人文的概念一经提出,就在世界范围内掀起了学术研究的热潮,出现了大批研究者、研究机构并孵化出诸多研究项目和技术成果。

1　研究背景和问题的提出

自2011年武汉大学数字人文中心成立以来,国内相继成立了多个数字人文研究中心,数字人文研究活动进行得如火如荼。同时,国内学者也开始关注国外相关研究进展,分析研究国外数字人文项目案例,总结项目特征和经验启示,以期为我国数字人文的发展提供借鉴。如林泽斐[2]对英国数字人文项目和特征进行了深入研究,为我国数字人文项目建设提供了宝贵经验;朱华顺[3]通过介绍美国图书馆数字人文实践案例,为我国图书馆开展数字人文服务带来启示;高胜寒等[4]对国内外数字人文研究成果进行文献计量分析,对比国内外研究发展状况,总结出数字人文领域的研究热点和发展趋势。

然而目前国内学者的关注点主要在于欧美国家,对于中国一衣带水的邻邦——日本的关注相对比较缺乏。笔者在CNKI中以检索式“SU=

* 本文系2020年度山西省回国留学人员科研资助项目“基于数字人文的山西省城市记忆资源开发策略研究”(项目编号:2020-098)的研究成果之一。

DOI:10.13666/j.cnki.jnlc.2021.0309

‘日本’AND SU=‘数字人文’”进行文献检索,仅得出15条结果(截止到2020年12月9日)。通过甄别后发现涉及日本数字人文项目的文献仅有1篇[5],且仅对日本立命馆大学艺术研究中心的数字人文实践进行研究,无法代表日本数字人文整体发展情况。自2011年日本数字人文协会(Japanese Association for Digital Humanities, JADH)[6]成立以来,日本数字人文研究迅速发展,随着研究团队不断扩大、研究项目不断增多,日本在数字人文研究方面取得了令人瞩目的进展和成果。由于日本与中国在地理位置、人文环境等方面的相近性,中日两国在政治、经济、文化、学术教育等方面的交流和相互影响日益扩大,因此有必要对日本数字人文发展情况进行分析总结,从而为我国数字人文研究提供经验启示。

2 调查概况

2.1 调查方法和数据来源

本研究通过对国际数字人文组织联盟网站centerNET上注册的日本数字人文中心以及日本数字人文协会(JADH)官方网站的分析,了解到日本数字人文研究主体以学术研究机构和各大高校为主。为全面把握日本数字人文项目研究进展,笔者从多渠道获取日本数字人文项目信息:对于研究机构,通过网络调查、文献研究及邮件咨询等方式获取研究项目信息,经过筛选后得到44条项目数据。对于高校,在科学研究费助成事业数据库[7](以下简称Kaken数据库)中检索有关数字人文研究信息。Kaken数据库是由日本文部科学省和日本学术振兴会共同资助的以高校为主体的学术研究项目信息公开数据库,是日本最权威的学术研究信息数据库。笔者以“Digital Humanities”“デジタルヒューマニティーズ”“デジタルアーカイブ”“人文情報学”四种表达式分别对数据库进行关键词检索,经过主题筛选和删除重复项后共得到954条数字人文项目相关研究数据(截止到2020年12月9日)。表1列出了Kaken数据库中主导数字人文项目数最多的前5名高校,表2是日本比较有代表性的数字人文研究机构和学会。

表1 日本高校主导数字人文项目数量排名TOP 5

高校	主导数字人文项目数量
东京大学	85
立命馆大学	81
筑波大学	34
京都大学	30
大阪大学	21

表2 日本数字人文研究代表机构

研究机构/学会	网址
国立国会图书馆下一代实验室	https://lab.ndl.go.jp/service/tsugidigi/
国立情报学研究所	http://www.nii.ac.jp/about/access/
人文学开放数据中心	http://codh.rois.ac.jp/index.html.en
国文学研究资料馆	https://www.nijl.ac.jp
国立历史民俗博物馆	https://www.rekihaku.ac.jp/english/index.html

2.2 日本数字人文整体情况概述

日本对人文计算的早期探索可以追溯到1966年,国立国语研究所使用大型计算机进行日语研究。1989年,日本信息处理学会成立“人文

科学与计算机研究会”(SIG Computers and the Humanities)[8],利用信息技术推进人文学的研究。进入21世纪,以国立国会图书馆为首的各种公共机构及高校相继开始了资料数字化的项目实践[9]。2002年日本文部科学省开始实施“21世纪COE计划”[10],COE是Center Of Excellence的缩写,意为“卓越研究中心”。该计划主要是对日本国内大学的优势尖端学科进行重点资助,培育世界领先的创造性杰出人才,推进具有国际竞争力的大学建设,目标是建立世界最高水平的研究基地。在该计划经费的资助下,日本国内各大学开展了许多大规模的项目研究,其中数字人文领域的研究也开始兴起并取得了一定的成果,培育了一批数字人文相关领域的研究者,如东京大学人文社会系教授下田正弘[11]、人文情报学研究所主席研究员永崎研宣[12]、国立情报学研究所教授北本朝展[13]等。进入2010年代数字人文研究进入了更加深入的阶段,2011年成立了日本数字人文协会(JADH),为加深国际交流合作,2012年JADH加入国际数字人文组织联盟(ADHO)。该协会每年召开一次数字人文大会,2020年度大会由大阪大学言语文化研究科主办,以“数字学术新的十年:缩影与核心”(A New Decade in Digital Scholarship: Microcosms and Hubs)[14]为主题。日本数字人文协会还创建了《人文情报学月报》(*Digital Humanities Monthly*)[15]和《日本数字人文期刊》(*Japanese Journal of Digital Humanities*)[16]等专业数字人文学术期刊。截止到2020年12月9日,日本共有三家数字人文中心在国际数字人文组织联盟网站centerNET上注册(见表3)。

表3　在centerNet注册的日本数字人文中心

数字人文中心	归属单位	成立时间
日本艺术文化数字人文中心(Digital Humanities Center for Japanese Arts and Culture)	立命馆大学	1998
下一代人文学开发中心萌芽部门(Center for Evolving Humanities)	东京大学	2007
人文情报学研究所(International Institute for Digital Humanities)	人文情报学研究所	2010

随着数字人文在日本的蓬勃发展,除了注册于centerNET的三家数字人文中心之外,越来越多的公共机构及高校也参与到数字人文项目的研究中。很多机构虽然没有冠以“数字人文中心”的名称,但实际也在积极推进数字人文相关研究。本研究将从日本各高校和研究机构主导的数字人文项目出发,探析日本数字人文的发展情况。

3　日本数字人文项目分析

3.1　学科领域分析

针对2.1部分获取的998条项目资料,笔者按项目所属的学科领域对数据进行补全整合,借鉴Kaken数据库对项目研究领域的分类,总结日本数字人文项目的主要学科领域分布情况(见图1)。由图1可见,日本数字人文项目研究涉及的学科领域相当广泛,其中图书馆情报学/博物馆研究、历史文化、文学艺术是日本数字人文最热门的研究领域。此外,教育学、天文地理气象等学科也是日本数字人文研究的热门方向。

3.2　研究内容分析

为深入了解日本数字人文项目涉及的具体内容,笔者对核心组织机构的项目进行了定性和定量分析研究,归纳出日本数字人文项目主要包含以下四个类别。

3.2.1　文献与档案研究

文献与档案研究是数字人文最常见的研究项目,也是数字人文研究的基础项目,主要通过数字化、文本挖掘等技术对传统纸质文献档案进行

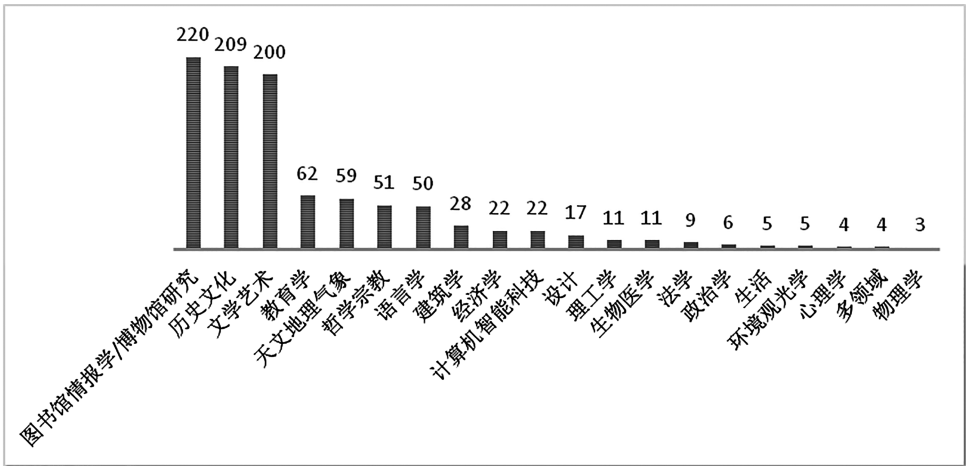

图1　日本数字人文研究学科领域分布情况

数字化保存和内容分析等。日本在文献档案方面的数字人文研究主要包括以下两方面：

(1)文献档案的数字化项目。出于永久保存和利用文献档案资源以及促进学术研究等的需要,包括图书馆、档案馆、博物馆等在内的多种组织机构对传统馆藏的纸质文献、古籍、图像档案等资源进行数字化存档和研究工作。纵观日本数字人文项目的整体情况,其中很大一部分的项目都是对不同历史时期各种类型的文献档案进行数字化存档,构建电子资源数据集。如日本国立情报学研究所开展的“东洋文库所藏稀有书籍数字化档案”项目[17],将与丝绸之路有关的稀有书籍、期刊论文等学术资料以统一的格式数字化并构建数据库,旨在建立一个使所有人都可以通过网络查看稀有书籍等文献的系统。此外,日本国文学研究资料馆实施的“日本历史典籍 Network 构建计划”(Project to Build an International Collaborative Research Network for Pre-modern Japanese Texts)项目[18],与国内外知名大学合作创建“新日本古典籍综合数据库”,将约30万本日本历史典籍拍摄成高清图像以 JPEG 的格式保存,并与现存的书目信息数据库集成,旨在建立一个日本历史典籍国际合作研究网络。

(2)文献内容和作者研究。对文献档案进行数字人文研究的另一个重要形式就是利用文本挖掘(Text Mining)、远读(Distant Reading)、可视化(Visualization)等技术,对特定文献的内容进行深入研究,以加深对文学作品或作者风格的理解。如大阪大学言语文化研究科的“狄更斯期刊作者分析”(Authorship Attribution and Stylometric Analysis of Dickens's Journals)项目[19],对英国作家狄更斯创建的《家常话》(*Household Words*)和《一年四季》(*All the Year Round*)两份周刊中狄更斯和威尔基·柯林斯的合著作品进行作者风格推定和计量文体分析。利用统计学的文体分析算法、文本挖掘方法及可视化方法进行数据分析,来推定合著作品中狄更斯的贡献度以及发生作者更替的具体位置,明确合著者的贡献程度,并开发出一种针对合著作品的高度精确的风格变化检测系统。

3.2.2　数字历史研究

随着信息技术的发展,从20世纪80年代开始,西方史学界开始兴起“数字历史”的潮流[20]。由上述学科领域分布特征可以看出,历史文化研究项目在日本数字人文项目中占有很大的比例。从研究目的来看,历史文化项目的研究重点在于历史档案的数字化存储,保护国内外有形和无形的历史文化遗产。笔者将日本数字人文项目中有关历史文化方面的研究归纳为三个方面:历史资料库建设、历史场景视觉重建和历史数据分析与可视化。

(1)历史资料库建设。通过建立历史资料库,以数字形式保存珍贵史料、社会记忆以及反映人类文明的各种形式的历史文化财产,便于人们了解、回顾重大历史事件。具体形式包括:(一)构建史料数据库以表现特定历史时期的鲜明特征。将古籍、图像等珍贵史料以数字形式保存在史料库中,借此向后人展现那些"消失的时代和消失的风景"。如日本国立情报学研究所和多所大学共同构建的华北交通档案(North China Railway Archive)项目[21],该项目公布了日军侵华期间刊登于报章杂志上的近40,000张华北交通写真,用户可通过多种检索方式进行检索查看,了解战争背后的故事以及当时中国的风土人情。(二)基于音频视频资源的口述史数据库。如上智大学的"来自东北的声音:311口述叙事档案"(Voices from Tohoku: 3.11 Oral Narrative Archive)项目[22],历时4年对2011年东日本大地震不同地区受害者进行采访,包括受灾情况、震灾造成的影响及灾后重建情况等,建立了总计超过200小时的视频数据库,并通过完全开放的社区网站"来自东北的声音"[23]提供给更广泛的受众。

(2)历史场景视觉重建。该类别的研究项目主要是通过地理信息系统(GIS)、人工智能(AI)、3D扫描、虚拟/增强现实(VR/AR)等技术,对历史场景、历史文化遗产等进行视觉重建,让用户身临其境地感受历史场景及文化活动。如人文开放数据中心(CODH)的"Edo+150项目"[24],利用最新的人工智能(AI)技术从江户时代的古地图中提取地名创建"江户地图β版",与现代地图信息相结合重现江户时代的空间信息。该项目还对江户时代出版的《江户购物指南》和《江户观光指南》中的关键信息进行可视化提取,与"江户地图β版"相关联,重建江户时代的商业和旅游业的空间信息。立命馆大学的研究者利用VR/AR技术重建京都的地理环境、街道和建筑等虚拟场景,通过点群计算开发日本重要节庆活动祇园祭山鉾巡行3D动画,结合各种历史图片、文本、音视频资料、3D地图等最大程度地还原祇园祭的真实场景[25]。

(3)历史数据的分析与可视化展现。此类项目主要对历史数字档案进行数据挖掘和数据分析,并将结果以可视化的方式展现出来。如日本人文开放数据中心的"武鉴全集"项目[26],利用差读(Differential Reading)技术对古典数据集《宽政武鉴》进行全文分析,建立关于江户时代大名家和幕府官员的个人信息和供职地理信息的核心信息平台。以其中的"参勤交代"部分为例,就是将不同时期官员轮流至幕府供职的情况以动画的形式在地图上进行可视化展现。

3.2.3 艺术和语言研究

艺术作为人文学科的代表领域,是日本数字人文研究的重点方向。对艺术领域的数字化研究给传统艺术带来了新的活力,便于人们更深刻地理解艺术所传达出来的内涵和文化。艺术项目研究涵盖了音乐、舞蹈、美术、电影等多个领域数字艺术档案的建设以及语言研究和语种保护等。

(1)数字艺术档案库建设。日本立命馆大学艺术研究中心[27](以下简称ARC)从成立之初就致力于为艺术领域的数字人文研究提供所需的数字环境和基础设施,使研究人员从事艺术文化方面的数字人文研究,在学术层面促进日本艺术文化信息的数字档案化事业。构建数字艺术档案不仅有利于珍贵艺术形式的数字化保存,也给科学研究提供了丰富的资源,为人文科学的研究提供了极大的便利,促进人文科学的研究向更广泛、更深入、技术更先进的层面发展。截止到2021年1月15日,ARC官网收录的由该中心独立或参与创建、运营的日本艺术文化相关的各类别的数据库总计超过80个,覆盖主题包括了日本传统歌舞伎、浮世绘、工艺美术、电影、演剧、摄影等。ARC还与加州大学伯克利分校东亚图书馆(C. V. Starr East Asian Library)合作,将该馆所藏的与日本相关的资料数字化,并建立数据库门户网站向公众开放[28]。该门户网站收录了江户时代的古典籍、日本制铜版雕刻、双六(日本古代一种棋盘类游戏)资料以及部分馆藏美术品拍卖目录。日本国立音乐大学附属图书馆与ARC合作将本馆所藏的《竹内道敬文库》资料进行数字化存档,建立"竹内道敬文库的世界"数据库[29],可提供三味线音乐演出相关的

浮世绘作品、音乐正本、演出票房排行和其他相关古籍的在线展览和检索功能,是日本音乐研究的重要资源。

(2)利用数字技术进行艺术研究。此类项目主要是在艺术研究中引入数字技术,深入挖掘艺术品背后的深层次语义信息,或在信息技术的辅助下,让传统艺术形式焕发出全新的光彩。ARC的“古乐器数字音色复原及数字演奏支援”项目[30],使用声线跟踪法和时域有限差分法(Finite-Difference Time-Domain,FDTD)等数字技术对19世纪末的古乐器实现音色复原,再现古音乐作曲时的音色效果,开发古乐器交互演奏的新界面,为古乐器创造出新的价值。为恢复濒临劣化、消失的文化遗产,东京艺术大学利用高精度三次元测量、高清数字图像处理和模拟方法开发出质量接近于原始图像的油彩画复制技术[31],将油彩画及雕塑作品中的画材成分、凹凸和笔触等准确复制,被称为“文化遗产克隆”[32]。截至2019年该项目结项,东京艺术大学的研究团队已成功复制出马奈的《吹笛子的少年》、梵高的《奥微尔教堂》和《自画像》等世界名作,并将勃鲁盖尔一世的《巴别塔》做成了立体像,为世界文化遗产的复原、保护和传承做出了巨大的贡献。

(3)语言研究和语种保护。对语言学领域研究的主要形式一般有大型语料库和数据集的建设与利用、语言文学作品的研究以及濒危语种或地方方言的保护等。如大阪大学言语文化研究科通过对赵元任、胡适、徐志摩等名家不同版本的翻译作品的内容对比分析,研究清末民国初期中国白话语言风格中的欧化语法现象[33]。北海道大学文学研究科基于TEI(Text Encoding Initiative,文本编码倡议)模型,使用Unicode、IDS(Ideographic Description Sequence)对日本平安时代编撰的字书《篆隶万象名义》《新撰字镜》和《类聚名义抄》进行技术处理,建立并发布平安时代汉字字书综合数据库(Integrated Database of Hanzi Dictionaries in Early Japan,HDIC)[34]。在濒危语种保护方面,东京外国语大学亚洲&非洲语言文化研究所使用田野调查法对缅甸少数民族克钦族的濒危语言进行收集、翻译、整理后以口述民族志的形式发布,并将数据在国际数字档案中注册保存,以此来保护传承传统文化和濒危语种[35]。

3.2.4 数字人文平台工具的开发

数字人文项目大多是以各种类型的图像和文本数据为对象进行的研究,与之配套的工具主要包括图像查看编辑分析工具、文本编辑识别与可视化工具等,日本数字人文工具开发研究主要包括发布文本或图像编辑工具、构建数字人文平台等。

(1)面向高精度图像查看编辑的数字化工具。IIIF(International Image Interoperability Framework,国际图像互操作框架)是一组实现图像资源互操作的技术标准,能使用户更方便地查看、编辑、比较和注释网络上的数字图像。为促进国际共同研究及共享研究成果,人文开放数据中心发布了符合IIIF标准的图像查看器IIIF Curation Viewer[36],不同于传统的图像查看器,该查看器具有独特的“策展”(Curation)功能,可对图像进行剪切、分类、收集和重新排序,并且可以按任何顺序查看。目前比较成熟的案例是“面部收藏”项目[37],通过裁剪和收集艺术品中出现的面部图像,比较面部的画法特征来研究不同作者与工作室之间的差异。随着研究的深入,人文开放数据中心基于IIIF Curation Viewer又开发出一系列配套工具,可对策展作品进行检索、编辑、管理和播放等,并构建IIIF策展平台IIIF Curation Platform[38],旨在通过支持从策展产生到利用的完整的工作流程来提高策展的多重利用性。

(2)面向文本的数字人文工具。面向文本的数字人文工具以OCR文字识别功能为主,如人文开放数据中心开发的KuroNet草书识别服务(AI OCR)[39]和KuroNet Text Editor[40]。KuroNet草书识别服务基于服务器API,可为符合IIIF标准的图像提供古籍中草书字体的识别及OCR功能。KuroNet Text Editor可以对识别出的结果进行编辑,调整识别字符的阅读顺序。该工具以IIIF Curation Viewer为基础,并且具有专门用于OCR文本编辑的添加和删除功能。

(3)数字人文信息发布平台。为方便研究者和对数字人文感兴趣的相关人员能及时全面地获取到数字人文研究的相关技术工具等资源,促进数字人文研究方法的开放化,关西大学亚洲开放化研究中心(Kansai University Open Research Center for Asian Studies,以下简称KU-ORCAS)[41]建立了东亚数字人文门户[42]。该门户将欧美数字历史公开研究项目The Programming Historian中采用的相关技术等信息资源翻译成日语公开发布,详细介绍了欧美数字历史研究中数字提取、数据清洗、网络可视化等最新技术;提供数字人文工具、网络资源和数字档案的发布平台,任何研究者只要拥有文献管理工具Zotero账号即可在平台上发布、更新相关资源;提供日语版TEI指南资源;汇总其他机构和项目发布的数字人文学习相关资源,如Python学习资源、TEI入门学习网络资源、数字人文相关书籍介绍以及KU-ORCAS数字化指南等。

(4)地理信息处理系统。日本国立情报学研究所开发的GeoNLP项目[43],旨在构建一个可提取自然语言句子中包含的场所描述(如场所名称和地址)并展示相关位置的地理标记(Geo-tagging)系统。该系统将地理信息科学(Geographic Information Science)与自然语言处理技术(Natural Language Processing)相结合,可实现在文章中出现的地名中嵌入经度和纬度等地理信息、通过周围环境识别具有相同符号的地名、为地名分配唯一ID,以及通过ID搜索包含相同地名的句子、识别句子中包含的地址并确定详细位置等功能。该系统已成功运用在如“2011东日本大地震新闻分析网站”[44]等多个网站中,用于从东日本大地震相关新闻中提取地名,并汇总县和地方政府或学校等公共机构的报道,以保存该场地震的相关记忆。

4 日本数字人文项目特征分析

数字人文作为一个跨学科研究领域,从诞生之日起就带着跨学科交融的天然属性。数字人文研究最初主要集中在文学和语言学领域,然而随着研究的深入和科学技术的不断发展,逐渐向更广泛的领域发展,覆盖了历史文化、艺术、地理、天文气象、医学、教育学等诸多领域。数字人文研究团队的成员通常来自不同领域,具备不同的专业背景。日本的数字人文研究除了具备数字人文研究跨学科交融这一基础特性之外,在数据开放、交流合作、研究重点和研究模式等方面也呈现出一些独有的特征。

4.1 数据开放化程度高

开放数据(Open Data)是指在互联网上发布的,任何人都可以获取、自由修改和再发布的数据[45]。利用相关技术进行数据搜集、处理并转换成结构化数据后,通过开放接口,以符合RDF(Resource Description Framework,资源描述框架)标准的格式为用户开放。目前各国在行政机构、文化学术机构等领域,特别是数字资源方面,只要没有特殊的法律限制,将资料信息作为开放资源公开的做法正变得越来越普遍[45]。日本数字人文项目研究的开放化程度非常高,从研究资源的获取、研究团队的组成、研究方法的采用到研究成果的发布,每一个环节都在积极推进数据的开放化进程。目前日本数字人文研究的开放化工作以数字资源的开放获取(Open Access)为主,其中比较大型的开放数据包括日本国立国会图书馆数字馆藏[46]、国立博物馆ColBase[47]、国文学研究资料馆“历史典籍Network事业开放数据集”[18]、国立历史民俗博物馆Khirin[48]、人文开放数据中心数据集[49]等。机构研发的数字人文工具和软件以及文本资料和图像数据集等基本都以OA的形式在Github平台上进行发布,促进软件的性能提升、增加使用者之间的良性循环。此外,IIIF[50]和TEI[51]等国际标准的积极推行和运用也体现了研究成果共享的开放理念。但开放并不意味着无限制地随意利用,日本在推进数据开放化进程的同时也没有忽视对数字资源和研究成果的产权保护。比如,针对一次信息,对于没有使用公费的研究项目的数据一般来说是不公开的,即便是公开也会附加严格的使用限制;对于二次资料,也

很少有无偿公开的情况,即使是无偿公开也会附以“仅限学术使用”的条件[9];有些开放数据还需要使用政府或相关组织团体指定的标准许可证或 Creative Commons(CC),对不同类型的资源附加不同的限制条件,确保开放数据可以在特定的限制范围内被合理合法地使用。

4.2 与国外研究机构交流密切

通过多种途径不断吸取国外研究经验,以促进数字人文事业的不断发展。具体来说主要包括:与国外图书馆、博物馆、展览馆等机构合作,将国外馆藏中关于日本的资料进行数字化保存。仅在 2020 年,ARC 就先后开展多个项目对印度萨拉加格博物馆、美国俄勒冈大学图书馆、斯坦福大学东亚图书馆、大英博物馆以及美国大都会艺术博物馆等机构馆藏中有关日本的历史文化等资料进行数字化存档。此外,采用国际通用标准(如 IIIF 和 TEI 等)进行图像和文本的编辑查看等工作,以推进国际共同合作研究和成果共享。积极参加国际数字人文交流研讨会、日本数字人文年度大会和其他各种形式的学术会议,也会邀请不同国家的学者进行演讲和项目介绍。通过参加国际性的学术会议等国际交流活动,及时掌握数字人文发展的最新动态信息,学习并使用国际先进的技术和研究成果。

4.3 注重亚洲历史文化传承

亚洲在人类文明发展史上有着举足轻重的地位,亚洲文化是文化研究的重要一环,是国际学术研究重点关注的对象。亚洲古文明对日本文化影响尤为深远,日本对亚洲历史文化的研究,不单是因为亚洲文化本身的古老神秘,还出于对本国文化源头的追溯[52]。随着计算机技术的飞速发展,研究人员越来越多地利用计算机技术推进本国和亚洲其他地域文化的探索、保护和传承。日本内阁府在《2018 年经济财政运营和改革的基本方针》[53]中强调,要将文化政策推进放在重点位置,构建国家文化品牌战略,提升文化产业的规模经济的同时,扩大文化遗产的高精细复制品制作和 VR 等尖端技术的应用。同时推进文化资源的数字化存档,实现“文化艺术立国”方针。在此背景下,日本各机构积极开展文化数字档案推进活动,其中规模最大的就是国立国会图书馆开发的文化资源汇总检索门户网站“Japan Search”[54]。此外,国立情报学研究所的“数字丝绸之路”(Digital Silk Road Project)项目[55]旨在利用信息技术收集丝绸之路相关珍贵史料、探索西域文明;东京大学人文社会系研究科的“佛教学新知识基础构建”项目[56],建立了超一亿字符的《大正新修大藏经》文本数据库[57]和图像数据库[58]等来保存和研究佛教藏经文化。

4.4 对研究成果进行深层次挖掘

从数字人文项目类型来看,大致可归类为数字化项目、文本挖掘图像分析项目、可视化项目等。纵观日本数字人文项目不难发现,这些类型的项目都不是互相独立的个体,项目之间存在着很强的关联性和继承性。以数字化项目为例,数字化项目的完成并不意味着研究的结束,以数字形式将资源存档也不是研究的终极目的,相比于资源的保存,研究人员更重视资源的灵活利用。将数字资源用作后续研究的数据来源和研究基础,对上一阶段的研究成果进行更深入地挖掘,在项目的层层推进中,促进更多新知识的发现和新技术的产生。如立命馆大学文化研究中心的学者利用最新的 3D 测量技术对位于印度尼西亚的联合国教科文组织世界文化遗产婆罗浮屠寺遗址进行精确数字存档[59],并以此数字档案作为实验数据开发出不受测量噪声影响的新型高精细可视化方法,今后还将利用该可视化方法开发遗迹的高精细“四维可视化内容”。日本数字人文项目在实践的过程中,不断挖掘项目深层次的研究价值和空间,扩展研究深度和广度。

5 对我国数字人文研究的启示

5.1 加大支持力度

科学研究离不开各个层面的支持。研究资源、科研团队、基础设施等每个环节都关系到研究成果的质量。本文认为主要应以下三个方面加大

对数字人文研究的支持力度。

(1)完善基础设施建设。基础设施是开展科学研究不可或缺的重要因素,完善数字人文基础设施建设需要从国家层面完善顶层设计,制定数字人文发展规划以及出台相关的政策措施扶持数字人文研究。为数字人文研究提供专项资金,在软硬件资源设备、工作平台、技术应用、活动经费等方面切实保障数字人文研究活动的顺利开展。此外,应积极发挥图书馆对于数字人文活动的支持作用,为数字人文研究提供数字资源、数据管理、数字人文馆员培训、信息技术、研究工具、成果发布等完善的配套支持。

(2)开展数字人文教育,培育数字人文专业人才。目前我国数字人文研究团队主要是由图书情报、计算机等专业的研究人员构成,缺乏数字人文专业人才。随着信息技术的发展,数字人文正成为当下学界热门的研究方向,"大数据视域下的数字人文研究"入选2018年度"中国十大学术热点"[60],高校应抓住这一重要契机,顺应时代潮流,大力发展数字人文教育。开设基础理论知识、数据管理、数据挖掘、数字人文工具使用等数字人文相关课程,提供数字人文课程教育和学位教育,培育一批优秀的跨学科复合型人才。

(3)促进数据开放进程。在互联网世界,开放数据可以说是一种公共财产,研究者可以轻松方便地获取所需资料,自由地进行研究开发活动。数字人文是"数据驱动型"研究领域,数据正在取代传统文献成为研究探索的素材和媒介[61]。数据开放可以有效促进信息资源的二次利用,使数据信息创造出新的价值,从而促进研究的发展。上海图书馆从2016年开始举办开放数据竞赛,借此向全社会征集开放数据开发和应用的成功案例,推进图书馆开放数据进程[62]。这是国内开放数据和图书馆转型的一次创新实践,值得其他机构参考借鉴。我国开放数据进程起步较晚,与发达国家相比还存在较大差距,主要表现在缺乏数据量、缺乏技术支撑和完善的制度保障等[63]。因此应尽快出台数据开放发展框架,构建数字人文平台;图书馆等相关主体应加快馆藏资源及其他资料的数字化,加大大规模数据库建设的资金投入,为开放数据提供更丰富的数字资源;以本体、RDF和关联数据为技术框架,更好地实现开放数据的五星标准[64,65];制定有针对性的规章制度,为开放数据提供制度保障,定期对开放数据进行维护提高开放数据的安全性;从内容、平台到服务三个方面促进图书馆等主体在数据资源、成果发布等方面的数据开放进程,促进数字人文研究中的开放获取(Open Access)和知识共享(Creative Commons)。

5.2 重视学科交叉,扩大交流合作

随着技术的进步、学科的发展,不同学科之间互相渗透、相互融合的现象已经越来越普遍。数字人文是典型的交叉学科,在不同学科领域的互相碰撞中促进新知识的发现和新技术的发展。这就需要我们重视学科交叉,将数字人文方法论共同体(Methodological Commons)[66]的思想灵活运用到更多的传统学科中。大胆创新、勇于实践,构建跨学科的合作模式,促进新型学科的产生和发展。多机构协同合作是数字人文的典型特征之一,在重视学科交叉融合的同时,数字人文研究还应注重国内外合作研究和学术交流、树立全球视野、积极促进大型国际联合项目的生成和顺利运作;积极组织参与各种形式的数字人文相关研讨会,了解最先进的研究热点和学科前沿,借鉴国内外研究团队的成功经验;坚持"引进来"和"走出去"相结合的开放理念,鼓励研究者积极参与国内外各学术组织联盟的学术讲座等;开展研究团队成员交流访学活动,学习国内外数字人文研究的先进理念、模式和技术成果,以便在与国内外不同研究团队的交流合作中激发新的研究灵感。

5.3 促进文化传承,深化历史研究

文化是一个民族的血脉,是一个国家宝贵的精神财富,也是国家强盛的重要支撑。我们要高度重视历史文化遗产的保护和传承,鼓励数字人文研究团队积极开展历史文化遗产的数字化保存和研究事业。国内学者在历史文化方面的数字人文研究中已经积累了不少成功的实践经验,如

北京大学参与的国际联合项目“中国历代人物传记资料库”(CBDB)项目、武汉大学和浙江大学合建的“数字敦煌”项目、上海图书馆“家谱知识服务平台”项目、中南民族大学建设的“唐宋文学编年地图”项目等,为数字人文历史文化研究做出了良好的示范。我们应尽快出台和完善针对历史文化遗产的保护政策,提升面向历史文化遗产的数字化理论、方法和技术水平,促进数字人文对历史文化的研究迈向更高的平台;加快文化遗产整合,构建与欧洲文化图书馆(Europeana)[67]类似的数字文化平台,促进中华传统文化遗产的永久保存和利用。

5.4 拓宽研究范围,加深研究层次

日本数字人文项目繁多,仅 Kaken 数据库中可检索到的数字人文相关研究就有近千例,且涉及学科领域众多,包括文献档案、历史文化、地理、艺术、政治、计算机、生物医药等,其中不乏一些历时近 10 年还在继续研究的大型项目。近年来,国内越来越多的高校、图书馆等机构参与到数字人文研究中,相继成立数字人文中心,积极推动数字人文各项研究。但是目前我国数字人文方面的研究项目还比较少,研究主题相对单一,且缺乏对项目的深入挖掘。因此,要将信息技术应用到更多学术问题的研究之中,不断拓展数字人文研究范围,促进研究主题的多元化;充分挖掘项目深层次的价值信息,最大程度地挖掘数字人文研究的潜力;促使数字人文向结构化、可视化、拟实化和智慧化方向逐步发展[68]。

数字人文研究方兴未艾,世界各国学者对数字人文领域的研究热度不断攀升。我国的数字人文研究虽然起步较晚,但也正在努力追赶中。如何实现弯道超车、后来居上对国内研究者而言是一项重大的挑战。在经济全球化的大背景下,各国之间的交流合作变得越来越密切,对学术研究而言也是前所未有的机遇。本文通过介绍日本数字人文研究的整体发展情况和其中部分代表性项目,总结日本数字人文研究的典型特征,以期为我国数字人文的发展带来启示和思考。

参考文献

1 Kirschenbaum M G. What is Digital Humanities and What's it Doing in English departments?[J]. Ade bulletin,2010: 55-61.

2 林泽斐. 英国数字人文项目研究热点分析——基于 DHCommons 项目数据库的实证研究[J]. 情报资料工作,2018(1): 97-104.

3 朱华顺. 美国图书馆数字人文案例研究及启示——以布朗大学、纽约公共图书馆为例[J]. 国家图书馆学刊,2016(6): 58-63.

4 高胜寒,等. 国内外数字人文领域研究进展分析[J]. 图书馆杂志,2016(10): 9-18.

5 徐彤阳,王霞. 日本立命馆大学艺术研究中心数字人文服务实践与启示[J]. 情报探索,2020(11):57-64.

6 Japanese Association for Digital Humanities[EB/OL]. [2021-01-15]. https://www.jadh.org.

7 KakenGrants[EB/OL]. [2021-01-15]. https://kaken.nii.ac.jp.

8 人文科学とコンピュータ研究会[EB/OL]. [2021-01-15]. http://www.jinmoncom.jp.

9 永崎研宣. 人文学分野とサイバーインフラストラクチャ: デジタル・ヒューマニティーズにおける現状と課題 (<特集> e-Scienceとその周辺~ 現状とこれから~)[J]. 情報の科学と技術,2013(9): 369-376.

10 グローバルCOEプログラム[EB/OL]. [2021-04-28]. https://www.jsps.go.jp/j-globalcoe/.

11 ResearchMap. 下田正弘[EB/OL]. [2021-04-28]. https://researchmap.jp/mshimoda.

12 ResearchMap. 永崎研宣[EB/OL]. [2021-04-28]. https://researchmap.jp/knagasaki.

13 ResearchMap. 北本朝展[EB/OL]. [2021-04-28]. https://researchmap.jp/kitamoto/.

14 JADH2020 [EB/OL]. [2021-01-15]. https://jadh2020.lang.osaka-u.ac.jp.

15 人文情報学月報/Digital Humanities Monthly[EB/OL]. [2021-01-15]. https://www.dhii.jp/DHM/.

16 Japanese Journal of Digital Humanities[EB/

OL]. [2021-01-15]. https://www.jstage.jst.go.jp/browse/jadh/-char/en.

17 National Institute of Informatics. Digital Silk Road Project Digital Archive of Toyo Bunko Rare Books[EB/OL]. [2021-01-15]. http://dsr.nii.ac.jp/toyobunko/index.html.en.

18 National Institute of Japanese Literature. NIJL-NW project[EB/OL]. [2021-01-15]. https://www.nijl.ac.jp/pages/cijproject/en/.

19 Authorship Attribution and Stylometric Analysis of Dickens's Journals[EB/OL]. [2021-01-15]. https://kaken.nii.ac.jp/grant/KAKENHI-PROJECT-15K02600/.

20 牟振宇.数字历史的兴起:西方史学中的书写新趋势[J].史学理论研究,2015(3):74-81.

21 North China Railway Archive [EB/OL]. [2021-01-15]. http://codh.rois.ac.jp/north-china-railway/.

22 Voices from Tohoku: 3.11 Oral Narrative Archive [EB/OL]. [2021-01-15]. https://kaken.nii.ac.jp/grant/KAKENHI-PROJECT-26285116/.

23 Voices from Tohoku[EB/OL]. [2021-01-15]. https://tohokukaranokoe.org.

24 Center for Open Data in the Humanities. Edo+150 Project[EB/OL]. [2021-01-15]. http://codh.rois.ac.jp/edo+150/.

25 祇園祭デジタル・ミュージアム2020-祇園祭の過去・現在・未来-[EB/OL]. [2021-01-15]. https://www.arc.ritsumei.ac.jp/lib/vm/gionfestivalDM/.

26 Center for Open Data in the Humanities. 武鑑全集[EB/OL]. [2021-01-15]. http://codh.rois.ac.jp/bukan/.

27 The Art Research Center[EB/OL]. [2021-01-15]. https://www.arc.ritsumei.ac.jp.

28 Japanese Special Collections in the C. V. Starr East Asian Library. University of California. Berkeley[EB/OL]. [2021-01-15]. https://www.arc.ritsumei.ac.jp/lib/vm/UCB/.

29 国立音楽大学附属図書館竹内道敬文庫の世界[EB/OL]. [2021-01-15]. https://www.arc.ritsumei.ac.jp/lib/vm/kunitake/.

30 Acoustic Sound Reconstruction and Playing-Support with Traditional Instruments [EB/OL]. [2021-01-15]. https://www.arc.ritsumei.ac.jp/e/works/jadh.html.

31 Development of Reproduction Technique of Oil Painting Works Using a High-resolution Digital Image Data and Analog Technique[EB/OL]. [2021-01-15]. https://kaken.nii.ac.jp/grant/KAKAKEN-PROJECT-15K12835/.

32 能够触摸的名画和雕刻,克隆文化遗产[EB/OL]. [2021-04-29]. https://web-japan.org/trends/cn/tech-life/tec171116.html

33 清末民国初期成立の漢語白話文体に見られる欧化語法現象の実証研究[EB/OL]. [2021-01-15]. https://kaken.nii.ac.jp/grant/KAKENHI-PROJECT-17K02753/.

34 The Establishment of the Foundation Study of Hanzi Dictionaries in Early Japan[EB/OL]. [2021-01-15]. https://kaken.nii.ac.jp/grant/KAKENHI-PROJECT-16H03422/.

35 失われつつあるカチン族の言語文化の記録・保存・継承[EB/OL]. [2021-01-15]. https://kaken.nii.ac.jp/ja/grant/KAKENHI-PROJECT-20K13024/.

36 Center for Open Data in the Humanities. IIIF Curation Viewer[EB/OL]. [2021-01-15]. http://codh.rois.ac.jp/software/iiif-curation-viewer/.

37 IIIF Curation Viewer. 顔貌コレクション[EB/OL]. [2021-01-15]. http://codh.rois.ac.jp/face/.

38 IIIF Curation Platform[EB/OL]. [2021-01-15]. http://codh.rois.ac.jp/icp/.

39 Center for Open Data in the Humanities. くずし字 OCR(AIくずし字認識)[EB/OL]. [2021-01-15]. http://codh.rois.ac.jp/char-shape/OCR/.

40 Center for Open Data in the Humanities. KuroNet Text Editor[EB/OL]. [2021-01-15]. http://codh.rois.ac.jp/software/kuronet-text-editor/.

41 Kansai University Open Research Center for

Asian Studies[EB/OL].[2021-01-15].https://www.ku-orcas.kansai-u.ac.jp/symposium/.
42 東アジアDHポータル[EB/OL].[2021-01-15].https://www.dh.ku-orcas.kansai-u.ac.jp/.
43 GeoNLP Project[EB/OL].[2021-01-15].https://geonlp.ex.nii.ac.jp/docs/.
44 東日本大震災ニュース分析[EB/OL].[2021-01-15].http://agora.ex.nii.ac.jp/earthquake/201103-eastjapan/mass-media/.
45 後藤真.人文情報学と歴史学[M]//後藤真,橋本雄太编.歴史情報学の教科書:歴史のデータが世界をひらく.東京:文学通信,2019:11-22
46 National Diet Library Digital Collection[EB/OL].[2021-01-15].https://dl.ndl.go.jp.
47 国立博物館 ColBase[EB/OL].[2021-01-15].https://colbase.nich.go.jp/?locale=ja.
48 国立歴史民俗博物 Khirin[EB/OL].[2021-01-15].https://khirin-ld.rekihaku.ac.jp.
49 Center for Open Data in the Humanities. DataSets[EB/OL].[2021-01-15].http://codh.rois.ac.jp.
50 International Image Interoperability Framework[EB/OL].[2021-04-29].https://iiif.io/about/
51 Text Encoding Initiative[EB/OL].[2021-04-29].https://tei-c.org
52 黎跃进.日本"丝绸之路"热及其成因探析[J].北方工业大学学报,2018(3):39-43.
53 日本内閣府.经済財政運営と改革の基本方針2018[EB/OL].[2021-01-15].https://www5.cao.go.jp/keizai-shimon/kaigi/cabinet/2018/decision0615.html.
54 Japan Search[EB/OL].[2021-01-15].https://jpsearch.go.jp.
55 Digital Silk Road Project[EB/OL].[2021-01-15].http://dsr.nii.ac.jp/index.html.en.
56 仏教学新知識基盤の構築[EB/OL].[2021-01-15].https://21dzk.l.u-tokyo.ac.jp/kibans/index.html.
57 The SAT Daizōkyō Text Database[EB/OL].[2021-01-15].https://21dzk.l.u-tokyo.ac.jp/SAT/index.html.
58 SAT Taishōzō Image Database[EB/OL].[2021-01-15].https://dzkimgs.l.u-tokyo.ac.jp/SATi/images.php.
59 Art Research Center, Ritsumeikan University. Digital Archiving of Indonesian Cultural Heritage and Development of 4D High-Definition Visualization Contents(インドネシア歴史文化遺産のデジタルアーカイビングと高精細4次元可視化コンテンツの開発)[EB/OL].[2021-04-28].https://www.arc.ritsumei.ac.jp/e/report/2020a3/pc/.
60 2018年度中国十大学术热点[EB/OL].[2021-01-15].http://theory.people.com.cn/n1/2019/0111/c40531-30516220.html.
61 刘炜,等.面向人文研究的国家数据基础设施建设[J].中国图书馆学报,2016(5):29-39.
62 赵宇翔,等.创意类开放数据竞赛作品评价指标体系构建与测定——以数字人文项目为例[J].中国图书馆学报,2020(2):75-95.
63 濮静蓉,等.国内外开放数据的研究进展与述评[J].图书与情报,2017(6):124-132.
64 夏翠娟.中国历史地理数据在图书馆数字人文项目中的开放应用研究[J].中国图书馆学报,2017(2):40-53.
65 5 star opendata[EB/OL].[2021-01-15].http://5stardata.info/cn/.
66 刘炜,等.数字人文研究的图书馆学方法:书目控制与文献循证[J].大学图书馆学报,2018(5):116-123.
67 Europeana[EB/OL].[2021-01-15].https://www.europeana.eu/en.
68 刘炜,叶鹰.数字人文的技术体系与理论结构探讨[J].中国图书馆学报,2017(5):32-41.

(徐彤阳　副教授　山西财经大学信息学院,顾婷婷　山西财经大学信息学院2020级图书情报专业硕士研究生)

收稿日期:2021-01-21