

JIS

第2卷
2023
第6期

第2卷
2023
第6期

智能社会研究

Journal of Intelligent Society

中华人民共和国工业和信息化部主管

哈尔滨工程大学主办

智能社会研究

Journal of Intelligent Society

中华人民共和国工业和信息化部主管



杂志公众号二维码
官网网址 www.jis.ac.cn



ISSN 2097-2091

9 772097 209239

定价: 45.00 元

ZHINENG SHEHUI YANJIU

目 次

全国大数据与社会计算会议专题

数字政府建设中的跨部门数据共享

——博弈行为与制度共建 董昌其 米加宁(1)

突发事件中短视频账号引发的群体情绪传播效果研究

..... 卫青蓝 马菁菁 余孟君 张 远(29)

基于隐私计算理论的短视频平台隐私悖论路径研究

..... 欧 露 何 翼 秦林瑜 唐珍名(45)

中国碳交易政策的经济后果与作用机制

——一个文献综述 杨嘉琦 齐佳音(63)

基于博物馆的文物大数据建设研究

..... 王家立 彭 亮 钟 意 张晓霞 陈建春(93)

数字与模拟译文专题

“数字与模拟译文专题”导言 亚历山大·加洛韦 著 王立秋 译(110)

论数字海 莎拉·普尔西奥(113)

作为文化技艺的编码

——论“数字”始于书写交流电 伯恩哈德·西格特(143)

模拟的黄金时代 亚历山大·加洛韦(159)

研究报告

新媒体舆论生态下的次生舆情防范与应对研究

..... 刘彦飞 梅议文 袁浩然 毛博文(181)

书评

社交媒体内容创作者的劳动与未来

——评《(不能)用你的爱好赚钱:性别、社交媒体与野心劳动》

..... 张竹楠(193)

访谈

算法、技术和文化的交汇点

——专访英国社会学家戴维·比尔 戴维·比尔 徐振宇(203)

CONTENTS

SPECIAL SECTION: CHINA NATIONAL CONFERENCE ON BIG DATA & SOCIAL COMPUTING

Cross-Departmental Data Sharing in Digital Government Construction: Game Behavior and Institutional Co-Construction	Dong Changqi, Mi Jianing(1)
Study on the Propagation Effect of Group Emotions Caused by Short Video Accounts in E-emergencies	Wei Qinglan, Ma Jingjing, Yu Mengjun, Zhang Yuan(29)
Research on the Privacy Paradox Path of Short Video Platform Based on Privacy Computing Theory	Ou Lu, He Yi, Qin Linyu, Tang Zhenyu(45)
Economic Consequences and Mechanisms of China Carbon Emissions Trading Policies: A Literature Review	Yang Jiaqi, Qi Jiayin(63)
Research on the Construction of A Cultural Relics Large Database Based on Museums	Wang Jiali, Peng Liang, Zhong Yi, Zhang Xiaoxia, Chen Jianchun(93)

SPECIAL SECTION: THE DIGITAL AND THE ANALOG

Introduction to “Special Section: The Digital and the Analog”	written by A. Galloway; trans. by Wang Liquiu(110)
On the Digital Ocean	S. Pouciau(113)
Coding as Cultural Technique: On the Emergence of the Digital from Writing AC	B. Siegert(143)
Golden Age of Analog	A. Galloway(159)

RESEARCH REPORT

New Media Public Opinion Ecology: Study on Secondary Public Opinion Prevention and Response Liu Yanfei, Mei Yiwen, Yuan Haoran, Mao Bowen(181)

BOOK REVIEW

The Labor and the Future of Social Media Content Creators: Review of (*Not*) *Getting Paid to Do What You Love: Gender, Social Media, and Aspirational Work*
..... Zhang Zhu'nan(193)

INTERVIEW

The Intersection of Algorithms, Technology, and Culture: An Interview with British Sociologist David Beer D. Beer, Xu Zhenyu(203)

基于博物馆的文物大数据库建设研究

王家立 彭亮 钟意 张晓霞 陈建春*

摘要:近年来,随着大数据、人工智能的快速发展,信息技术在文博领域的应用日益广泛。为建立文物资源大数据库,进一步推动文物数字化保护、活化利用,提升中华文化的影响力,增强文化自信,本文依据博物馆行业现状,针对文物资源大数据库建设的路径和挑战,梳理了全国博物馆和文物状态现状、文物元数据标准建立以及文物数字化工作开展情况。本文提出,在细化文物类别数据标准和文物元数据标准体系的基础上,深化文物数字化身份识别码管理体系,探索建立区域性文物数据中心,实现博物馆行业文物资源大数据库及全国文物大数据交互平台;并就大数据在博物馆陈列展览、科学研究、公共教育三个方面的应用,分析其面临的挑战和解决思路。最后,本文对文物大数据的发展进行了展望。

关键词:文物数字化 大数据 博物馆 文物元数据 区域性文物数据中心

一、引言

《“十四五”文物保护和科技创新规划》指出:“‘十四五’时期是我国开启全面建设社会主义现代化国家新征程的第一个五年,也是推进社会主义文化强国建设、推动实现从文物资源大国向文物保护利用强国跨越的关键时期。”中华文化有着深厚的底蕴和丰富的文物资源,“十三五”期间,国家文物局牵头完成了第一次全国可移动文物普查,对全国文物保护保存状况进行了摸底。“十四五”规划明确提出要“建设国家文物资源大数据库”:“系统整合全国不可移动文物资源数据库、国有可移动文物普查数据库、革命文物数据库等,加强文物资源大数据应用。”“建立文物数字化标准规范体系,健全数据管理和开放共享机制,加

* 王家立,重庆红岩革命历史博物馆;彭亮(通讯作者,txxy25pl@sohu.com),重庆红岩革命历史博物馆;钟意,重庆红岩革命历史博物馆;张晓霞,重庆红岩革命历史博物馆;陈建春,重庆红岩革命历史博物馆。

大文物数据保护力度。”(国务院办公厅,2021)

国际博物馆协会(International Council of Museums, ICOM)对“博物馆”的最新定义翻译如下:“博物馆是为社会服务的非营利性永久机构,致力于研究、收集、保护、解释和展示物质和非物质的文化遗产。博物馆对公众开放,具有开放性和包容性,旨在促进多样性和可持续性。它们以道德和专业的方式运作和交流,并参与社区建设,提供多样化的教育、享受、思考和知识共享体验。”(ICOM, 2022)陈红京认为,博物馆是征集、保护、研究和传播文物的机构,而数字化博物馆则是实体博物馆在空间和时间上的延伸,具有诱发观众需求和辅助理解陈列展览等功能(陈红京,2019)。数字化博物馆的显著特征是以数字形式存储藏品信息,而这些信息源自文物的数字化。通过数字信息采集技术,文物的各类信息可以以数字形式保存和利用起来(刘毅,2019)。

随着人民群众日益增长的精神文化需求,中国博物馆的数量也在持续增加。截至2021年,中国大陆备案的博物馆总数为6183家(国家文物局,2021a)。^①与此同时,博物馆的文物数字化工作有序展开。但由于不同博物馆之间的技术能力和资源水平存在差异,各博物馆的文物数字化进程也呈现出差异化的情况。本文旨在思考文物资源大数据库建设的路径和挑战,并着眼于未来大数据在博物馆行业的可行应用,同时思考如何解决其所面临的挑战。笔者对全国博物馆行业和文物整体现状进行了梳理,并调查了文物元数据标准的建立以及文物数字化工作的开展情况。本文提出,在细化文物类别数据标准和文物元数据标准体系的基础上,深化文物数字化身份识别码管理体系,并探索构建区域性文物数据中心,逐步实现构建博物馆行业文物资源大数据库乃至全国文物大数据交互平台的目标。此外,本文还通过思考文物资源大数据在博物馆陈列展览、科学研究、公共教育等方面的应用延伸,探索解决数据安全、文物版权、人才培养等多个问题的策略,展望了文博领域大数据的发展前景。

^① 统计数据未包含香港、澳门、台湾的博物馆备案数量,下同。

二、博物馆数字化发展现状分析

(一) 博物馆建设和文物藏品情况

近 5 年来,全国博物馆的数量迅速增长,但是博物馆的质量等级和分布情况并不均衡。根据国家文物局(2021a)公开信息,2021 年全国备案的博物馆情况如图 1 所示。其中,已定级的博物馆有 1218 家,仅占全国博物馆总数的 19.70%。这意味着还有大量博物馆的硬件和软件配套设施仍然不完善。此外,各省市的博物馆数量分布也不平衡。例如,山东省的备案博物馆数量为 623 家,占全国备案总数的 10.08%;浙江省、河南省、广东省、四川省、江苏省和陕西省的备案博物馆数量均超过 300 家;相比之下,天津市、宁夏回族自治区、青海省、海南省和西藏自治区的备案博物馆数量都少于 100 家,其中西藏自治区的备案博物馆数量最少,仅为 12 家,占全国备案总数的 0.19%(国务院第一次全国可移动文物普查领导小组办公室、国家文物局,2017a)。

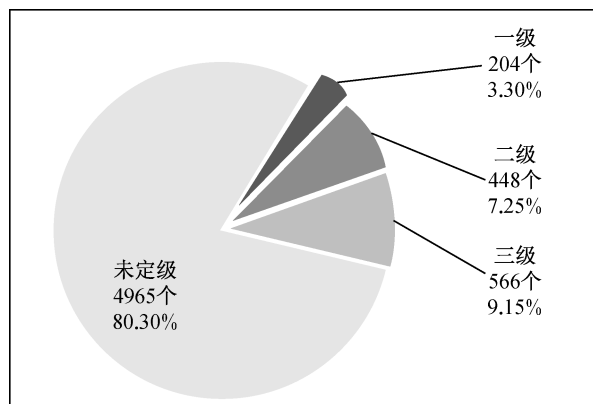


图 1 备案博物馆质量等级和数量分布百分比

根据《第一次全国可移动文物普查工作报告》,全国可移动文物总数为 6407 万件/套。其中,按照普查统一标准登录完整信息的文物数量为 2661 万件/套(41.53%)。所采集的文物基础信息包括 15 项内容,如类别、年代、级别(文物等级)、来源、入藏时间、完残程度、保存状态等(国务院第一次全国可

移动文物普查领导小组办公室、国家文物局,2017b)。总体而言,为了建立一个具有交换流通功能的全国性文物大数据平台,当前文物普查统一标准记录中的数据项还无法满足相关需求。因此,需要进一步完善和补充每件/套文物的数据内容。

(二) 博物馆文物元数据标准建立情况

文物元数据是进行文物数字化的基础,而其规范体系则是构建文物大数据平台的依据。《文物元数据体系(征求意见稿)》指出:“在面向数字化保护的标准规范体系和文物数字化生命周期模型的基础上,建立一整套具有可操作性、适用性和可扩展性的文物数字化元数据规范体系。”其中,明确将文物元数据类型分为:专门元数据、资源集合元数据、管理/保存元数据、服务元数据和元元数据(朱强、肖珑、张俊娥等,2017)。各类元数据类型整理如下(表 1):

表 1 文物元数据类型		
元数据分类	元数据类型	说明和举例
描述元数据	专门元数据	描述文物的特色内容、属性等
	资源集合元数据	基于文物实体的数字资源对象集合,包含主题、文物类型、文物年代等
针对数字资源管理	管理元数据	数字对象采集、加工中所涉及的数据
针对元数据保存	保存元数据	国际上建立了多个比较通用的保存元数据格式,如 NLA、NEDLIB、CEDARS、PREMIS
针对 Web 服务的元数据	服务元数据	规范文件格式、编码规则、统一标识等
元数据管理	元元数据	规范元数据的数据

文物元数据体系的建立,为文物数字化保护和利用奠定了坚实的基础。然而,目前的元数据标准并未针对革命文物建立专门的科目。由于国际标准对中国革命类文物的基本认知不足,常用的元数据标准主要关注物件本身的价值,而我国的革命文物大多数并不具备很高的经济价值,但其所涉及的历史事件和英雄人物却具有重要的研究价值和拓展价值(练洁、李娉、赵星宇,2021);因此,针对革命文物,同样需要制定符合其特点的元数据标准。此外,刘峤等人指出,在知识图谱构建技术中,信息抽取主要涉及实体抽取、关系抽取和属性抽取(刘

峤、李杨、段宏等,2016)。革命文物相关的佐证材料、视频说明、亲属访谈等内容数据,可以作为这三种技术的有效输入内容,其中的事件、人物、地点、时间、相关性文物资源归属等关键词可以直接作为信息抽取的结果。因此,在制定革命文物的元数据标准时,根据其区别于其他文物的特点,能够获取大量的关联信息,并根据这些关联信息对元数据指标进行设计,从而有效地促进后期对文物数字化数据的活化利用。

(三) 博物馆文物数字化工作开展情况

针对文物的数字化保护,毛芳提出了六条技术路线,包括制定采集规范及标准、全面获取可移动文物背景信息、利用高精度图像技术获取文物高保真图像、利用现代化技术获取文物本体信息、可移动文物数字化保护信息录入以及可移动文物数字化保护成果利用(毛芳,2014)。本文在考虑现有实际情况的基础上,对这些步骤进行了整理和补充,总结出以下六个步骤:

(1) 细化完善采集规范、元数据标准。目前施行的指标体系参考了《博物馆藏品信息指标体系规范(试行)》(国家文物局,2001b)的内容。然而,由于该指标体系制定的时间较早,硬件设备的采集能力和元数据标准在此期间已经有了长足发展,因此必须对该指标体系进行修订、细化和增补。

(2) 获取可移动文物具体信息。博物馆文物征集管理部门和研究部门需要按照《馆藏文物影像拍摄规范》《馆藏文物著录规范》和《博物馆藏品信息指标体系规范(试行)》的标准,收集文物相关的信息和资料,并进行整理和归纳。

(3) 获取文物高保真图像。朱强等人提出,文物的数字对象总体上可以分为文本对象、图像对象(二维图像、三维图像、动态图像)、音频对象、多媒体对象和复合对象(表2)(朱强、肖珑、张俊娥等,2017)。针对二维平面文物,通常采用高清影像拍摄和高分辨率扫描等技术获取高保真图像;针对三维文物,则采用360°、720°多角度环形高清影像拍摄,数字点云扫描等技术,结合三维仿真模型和图片拼接,形成文物的三维数字化数据;针对不可移动文物,常用的采集方式是三维图像扫描技术,以获取文物的高精度三维空间数据和色彩纹理信息,并利用三维模型还原文物本体。

表 2 文物数字对象		
文物数字对象		说明
文本对象		古籍、文档
图像对象	二维图像	记录文物的背景、出土情况、类型、尺寸等信息
	三维图像	记录文物的体积、外形、纹理、装饰、图案、颜色、光泽等信息
	动态图像	立体古建筑、雕塑、瓷器
音频对象		动态 gif 图像、流媒体
多媒体对象		配乐
复合对象		上述多种形式结合展示
		需结合文本、二维、三维等信息
		资源集合
		上述多种形式、多种对象的集合

(4) 获取文物本体信息。利用文物科技保护技术,如高光谱信息采集、超景深显微镜、纤维染色分析等技术手段,对文物本体的材质、病害情况进行采集,为文物的保护环境和修复工作提供基础信息。例如,可以使用便携式测色仪采集文物的颜色信息,采用荧光光谱仪采集金属元素信息,等等(毛芳,2014)。

(5) 进行数字化保护信息录入。完成文物数字化信息采集后,需要对其进行整理、归纳并存档。随着高清技术的发展,文物数字化保存的数据量级逐渐提升,包括高分辨率、超大幅面和无损的数据。数据保存涉及数据库系统、数据库物理环境和数据灾备策略等多个方面的建设,需要相应的技术和资金支持。以数据为例,第一次全国可移动文物普查工作总共登记了 5000 万张文物照片,数据总量超过 140 TB(国务院第一次全国可移动文物普查领导小组办公室、国家文物局,2017b)。其中,按普查统一标准登录文物完整信息的仅有 41.53%。根据目前的文物数字化元数据规范体系试行标准,需要记录的文物基础信息将会多于此次普查采集的 15 项信息。因此,未来文物数字化保存的数据量将更加庞大。

(6) 利用数字化保护成果。目前,主要采用光学字符识别技术(optical character recognition, OCR)进行平面可移动文物的活化,将平面图文内容转换为数字化内容(张当中,1997)。然而,可移动文物保护资金投入的重点仍在数

字化保存方向,对文物活化模型的设计和训练尚未在政策上明确资金支持,导致数据挖掘的深度较低,未能实现文物的深层次活化利用。

在文物数字化的实施过程中,数据研究、应用开发、展览展示和文创产出,受到人才、技术和资金等多种因素的影响,导致各博物馆在相关环节的工作开展进度不均衡。根据笔者了解到的行业信息,目前全国大部分博物馆并未全面实施文物数字化保护工作,只有少量数字化基础较好的省市级博物馆基本完成了文物数字化保护工作,并开始关注文物数据的活化利用和深度挖掘研究。

为了推动文物数字化保护和活化利用,建设国家文物资源大数据库,笔者认为应该以文物类别数据标准、文物元数据标准体系、文物数字化身份识别码管理体系以及区域性文物数据中心的建设为努力的重点。同时,通过分级权限管理和行业行政审批等手段,进一步建立和完善国家文物数字化管理体系。

三、博物馆馆藏文物数字化发展建议

(一) 完善标准体系

1. 完善文物类别数据标准和文物元数据标准体系

文物数字化工作的开展,基于文物类别数据标准和文物元数据标准体系的完善。文物分类是进行文物保护、文物研究和文物利用的前提。由于各国国情不同,文物类别的划分相对复杂,目前尚无国际通用标准。国际标准通常参考盖蒂研究所(The Getty Research Institute)发布的《艺术与建筑词汇表》(*Art & Architecture Thesaurus*, AAT),该词汇表通常用于描述与艺术、建筑和其他文化遗产项目相关的受控词汇(Getty, 2017)。

《中华人民共和国文物保护法》总体上将文物分为不可移动文物和可移动文物,第三条规定:“古文化遗址、古墓葬、古建筑、石窟寺、石刻、壁画、近代现代重要史迹和代表性建筑等不可移动文物,根据它们的历史、艺术、科学价值,可以分别确定为全国重点文物保护单位,省级文物保护单位,市、县级文物保护单位。历史上各时代重要实物、艺术品、文献、手稿、图书资料、代表性实物等可移动文物,分为珍贵文物和一般文物;珍贵文物分为一级文物、二级文物、三级文物。”

2017年,北京大学等7家机构结合传统分类属性和中国文物特色主题属

性,联合发布《文物分类体系(征求意见稿)》,形成了一部结构化词表,包括文物分类和文物描述相关的词汇(黄明玉、冯英、张俊娥等,2017)。

笔者认为,在文物类别数据标准基础上,仍需进一步完善文物元数据标准体系。应全面考虑革命文物的实际情况,在制定元数据标准体系过程中,根据革命文物的特点制定并完善革命文物元数据标准。

馆藏革命文物大多是通过革命历史人物、人物亲属或与革命历史事件有关的群众捐赠或征集而来的。由于革命文物的历史背景区别于其他文物,仍具备通过人事档案资料、现存佐证人等途径获取佐证材料证明的可能性。佐证材料所承载的信息、革命文物相关人物当事人或亲属的访谈记录、视频采访资料、文博研究馆员的研究等,都是革命文物背景信息的重要组成部分。革命文物信息除应记录藏品所在机构外,还应记录文物所经手的所有机构或所有人的历史记录。例如,重庆红岩革命历史博物馆收藏了国家一级文物“杨虎城出国纪念锦旗”,该文物的背景信息就记录了杨虎城出国后锦旗由其秘书王菊人收藏,后于1964年10月由王菊人捐赠给重庆歌乐山烈士陵园(红村网,2016)。部分革命文物虽然指向唯一历史人物,但由于历史原因可能收藏于不同的机构或个人,故其相关革命文物资源归属信息也应作为革命文物元数据的重要内容予以展示。例如,重庆红岩革命历史博物馆珍藏了江竹筠烈士的七封家书,这是她好友谭竹安捐赠;而她就义前的最后一封信,即她的遗书,则现存于重庆中国三峡博物馆(华龙网,2021;富治平,2007)。有效设置江竹筠信件的资源归属信息元数据,有助于研究人员查阅革命文物数据并开展科学研究工作。

本文对标艺术作品著录类目(Categories for the Description of Works of Art, CDWA),基于革命文物实际情况,新增了部分革命文物专有元数据(表3)(Bacca, Harpring & Ward et al., 2022)。

表 3 革命文物专有元数据		
CDWA 元数据	革命文物新增 专有元数据	说明
题词/标记 (inscriptions/ marks)	专家题词或评论	相关文博专家或权威人士的题词、批注等信息
保存/修复史 (conservation/ treatment history)	修复情况	革命文物的修复历史、修复人等信息

(续表)

CDWA 元数据	革命文物新增 专有元数据	说明
文本信息 (context) 相关文本参考信息 (related textual references)	佐证材料	革命文物历史背景的有关佐证材料信息
	关系人访谈资料	和革命文物有关的亲属、朋友、群众等的访谈信息,作为革命文物背景的重要补充
	证词证言	和革命文物有关的人物提供的证词、证言
现藏地点 (current location)	资源归属信息	革命文物,如书信、文稿等,可能分散收藏于不同的机构或个人,该信息应记录所有机构或所有人信息
所有/收藏历史 (ownership/collection history)	文物所有机构或个人的记录	革命文物历史所有机构或所有人的信息,移交、转赠等记录
版权/限制 (copyright/restrictions)	数字化数据及其衍生 IP 版权信息	应记录该革命文物数字化数据及其关联 IP 的版权归属信息
相关的可视资源 (related visual document)	视频资料	革命文物相关的视频影像资料可作为佐证材料,如当时拍摄的照片、影片等,记录了某位革命历史人物曾使用该物品

此外,随着高清拍摄和扫描技术的发展,文物数字化数据量也大量增长。笔者认为,不应盲目追求最大分辨率、更高清晰度来采集数据,应结合文物大数据建设中实际需要的算法模型、博物馆数据存储能力、文物大数据平台传输能力等因素,探索适当的数字化采集分辨率区间,避免成本浪费、技术冗余和传输困难。

2. 建设文物数字化身份识别码管理体系

国家文物局在完善文物标准体系的基础上,构建了文物数字化身份识别码及配套管理制度,在完成文物数字化保护工作后,依据元数据标准赋予其唯一的身份识别码。

笔者认为,该识别码应增加对应字段,用以记录文物的必要数字信息。例如,在文物类别、年代、文物等级、来源、入藏时间、完残程度、保存状态等 15 个基础字段之上,增加文物所属省份、所属博物馆、数据生成时间、录入人员等信息。每一件/套文物对应唯一身份识别码,该识别码上传至国家文物资源大数据库,纳入文物管理部门以便检索、管理。利用身份识别码对文物进行管理,能够更好地做到以下工作:

(1) 通过射频识别 (radio frequency identification, RFID) 等技术辅助,更好地实现对库内文物的准确快速清点、盘库、安全检查、存储囊匣位置定位等管理

操作。

(2) 文物本体在外出展览、修复或运输时,每一个经手单位通过扫描二维码等标签读取文物信息,实现时间戳信息的同步、记录,有利于文物管理部门系统化开展备案、存档、督查等工作。

(3) 依照相关管理模式,可依托文物管理部门对数字化的文物数据系统开展备案、审批和流转记录等工作。

通过以上模式,文物管理部门可以随时查看某文物本体或数字化数据所在位置或机构,并以精确到秒的时间查看文物流转各节点的信息;文物版权所在机构也可以随时追踪文物本体或数据的动向。

利用文物数字化身份识别码管理体系,可以有效地管理文物本体和文物数据,打击非法文物和文物数据贩卖,促进健康的文物数据共享交流和数据交易,有利于文物的活化利用。

3. 标准体系建设的建议

我国拥有大量博物馆,但其质量等级分布不均。在文物数字化标准体系的建设过程中,应充分考虑与现有《博物馆运行评估标准》等标准和体系对接,结合全国一级、二级、三级和未定级博物馆的综合实力,与博物馆的实际运行状况相协调,建立符合国情的博物馆和文物数字化标准体系。

(二) 培育区域性文物数据中心

《“十四五”文物保护和科技创新规划》提出,应建设国家文物资源大数据库,加强文物资源大数据应用。鉴于我国博物馆行业发展的不平衡性,在建立文物资源大数据库时,应充分考虑各个省、自治区、直辖市的博物馆实际情况。

在经济发达、博物馆和文物行业发展较好的地区,人才、资金等方面具有一定优势。这些地区拥有较多数量和较高质量的博物馆,其藏品资源和科研技术方面也会得到政策方面的重视和支持(李文昌,2008)。文物数字化涉及博物馆内部的文物征集、保护、管理和信息技术等多个部门的协作。然而,在全国备案博物馆中,只有不到20%的已定级博物馆在内部管理上符合《博物馆运行评估标准》;而未定级博物馆在人员、技术、资金等方面存在不足,在实现文物数字化方面面临一定的困难。因此,一些中小型博物馆迫切需要地方政府或该地区高质量博物馆在数字化建设中予以技术和人才支持。

应综合考虑博物馆行业发展不平衡的现状,结合文物数字化保护的实施步骤,建设国家文物资源大数据库。在文物信息采集阶段,使用高新技术进行采集工作需要大量人力、设备投入以及项目建设成本;而利用文物数字化保护成果,进一步进行数据挖掘和分析,构建更丰富的文物知识网络,如建立知识图谱,拓展文物信息存储、传输、共享和交易模式,更需要海量的计算机资源支持,包括算力、存储和运维等方面的资源(刘峤、李杨、段宏等,2016)。

通过对全国博物馆的发展和藏品数量状况进行分析,本文认为可以从各省市高质量综合博物馆和专业博物馆中,选取一批作为该区域的文物数字化保护中心,利用这些博物馆在资金申报和人才聚集方面的优势,为该区域内其他博物馆提供文物数字化保护服务。这样做可以降低其他博物馆购买设备或进行项目建设的成本,从而降低该区域内数字化保护工作的软硬件总投入。

鉴于文物数据的特殊性,《中华人民共和国文物保护法》第五条明确规定:“属于国家所有的可移动文物的所有权不因其保管、收藏单位的终止或者变更而改变。”但是,《中华人民共和国文物保护法》并未对馆藏文物数字化成果的调拨、交换、借用、复制、共享、交易等行为的可行性进行明确规定,因此文物数据的汇聚尚未获得法律法规确权。有鉴于此,可以基于区域文物数字化保护中心的搭建,利用 NoSQL 数据库、数据仓库、区块链等技术形成区域性文物数据中心,采用分布式存储技术解决文物数据保管归属的法条适用问题,通过配套建设省级和国家级审核节点,汇聚形成全国文物资源大数据库。与此同时,还可以利用区域性文物数据中心试行有关的文物数字化标准,在实践中不断总结完善相关标准,最终形成可全国执行的标准体系。

四、大数据在博物馆场景的应用思考

随着国家文物资源大数据库的建设,相关大数据应用必将为博物馆行业带来新的助力。目前,在博物馆行业中成功应用大数据的典型案例还比较少。笔者认为,博物馆行业的大数据应用应着眼于观众预约、数字导览、游客画像分析、展品位置导航、文物研究、陈列展览、研学教育、文创产品等博物馆业务,并与博物馆管理、运行评估等工作流程进行深度结合。邓慧超认为,大数据可应用于博物馆藏品信息资源共享、信息资源传播与传播数据良性循环、观众行为分析、管

理数据分析等方面(邓慧超,2018)。结合博物馆的基本业务,笔者从博物馆业务的三个方面思考了大数据的应用。

(一) 大数据在陈列展览中的应用

陈列展览作为博物馆的基本业务,具有广泛的大数据应用前景。在中国,“陈列”一般指长期稳定的展览,而“展览”则指更加灵活的短期展览,但实质上两者并无差异(王宏钧,2001)。本文将这两个概念视为同一概念,都指博物馆的展览。根据展陈时长,中国的博物馆陈列展览可分为基本陈列、专题陈列、临展和特展。除了本馆的基本陈列和专题陈列外,每个博物馆都会举办大量的临展和特展(龚青,2011)。博物馆自主策划的临展和特展,可以根据主题多样化、内容丰富度和展览深度等一系列指标进行考核。对于博物馆自主策划的展览,选取适宜的主题尤为重要,以吸引观众,迎合当前的知识需求热点,满足人民日益增长的精神文化需求。预训练了大量文博类数据的模型,如结合了大语言模型的微软新必应(New Bing)搜索引擎,具备强大的信息检索和整合能力,可以帮助策展人更快速、准确地确定展览主题和章节。在确定主题和章节之后,基于大数据的大语言模型还可以在收集查询展览文献资料、分类归档和编写展览大纲等环节发挥重要的辅助作用。

此外,现有陈列展览经常同步规划一系列同主题、同类型展出,而一些博物馆单靠自身力量难以完成这个任务。因此,借助大数据应用,在国家文物资源大数据库的支持下,以文物数据互通为基础,促进多个博物馆之间的合作办展将成为必然趋势。在国际交流合作中,考虑到当前的文物出入境管理政策,文物数据比文物本体更易于实现交流合作,能够促进国内外博物馆在学术科研、陈列展览方面的加速发展。

在布展时,可以利用大数据比对相似主题和同类展览,优化空间布局,甚至利用人工智能模型直接生成效果图,以便更直观地进行布展准备和实施工作。

在导游导览环节,大数据与人工智能的结合,可以为观众提供更优质的观览体验。虚拟智能讲解员利用聚类分析、用户画像分析等技术,可以对游客的游览行为和展品兴趣进行数据分析。相较于真人导游,智能讲解员更了解游客的心理需求和行为模式,能够提供针对性更强的个性化导览服务。它能为每位游客提供他们所希望了解的文物知识和图像细节,并能解答关于展品、历史和文化等

方面的问题。类似于新必应搜索引擎这样的应用具备强大的信息收集能力,可以为游客提供丰富的补充信息和背景知识,而不再局限于固定的路线讲解词。当行业大数据足够汇聚后,还可以形成更具个性化的展品推荐和线路,形成个人导游地图,帮助游客自主探索感兴趣的展区和展品。

(二) 大数据在科学研究中的应用

大数据的应用能够极大提升文博领域信息检索和整理归档的效率,同时有助于实现文物的全面活化。国家文物资源大数据库的建设,可以助力文博领域数据挖掘,并利用语义网络技术和知识图谱技术构建更全面的文物知识图谱;采用图形化的数据可视化展现形式,有助于文博人员进行学术研究(刘峤、李杨、段宏等,2016)。例如,当文博人员需要了解某件文物的更多背景信息时,借助知识图谱,在输入关键词后,大数据会自动匹配与该文物相关的属性、人物、地点、事件等信息,帮助文博人员快速完成信息收集;此外,结合数据相关性,还可以列举更多高相关性的信息,极大降低文博人员在信息检索上的难度。目前,在文博领域,已经涌现出许多依托大数据建立知识图谱的案例。例如,葛勇文利用有关“中共一大”文物的数据,通过相关数据处理和知识图谱构建技术,成功构建了相关的革命文物知识图谱。该图谱成功实现了革命类纪念馆之间数字资源的强关联融合,突破了地域和机构之间的阻隔,促进了革命历史文物知识的共享。这样,革命文物与历史之间建立了联系,挖掘出了革命文物背后的史实故事(葛勇文,2021)。

经过训练的自然语言处理(natural language processing, NLP)模型可以帮助文博人员进行文档资料的整理归档工作。以文献资料的分类整理工作为例:以前,文博人员需要花费较长时间来总结文献简介、整理出版信息,并为其赋予对应的类别标签;NLP模型可以实现分类整理任务的智能化,从而大大减轻文博人员的工作负担。此外,结合大数据和人工智能技术,对报刊、书信、档案等文字段落的信息提取、关系提取、主题提取等操作进行优化,是实现平面纸质文物活化利用的必经之路。

(三) 大数据在公共教育中的应用

在公共教育方面,将大数据技术和人工智能应用结合起来,可以生成各种教

育主题和内容,不断拓展教育的广度和形式。在这个基础上,还可以利用知识图谱技术进一步扩充学习内容的丰富度。例如,针对博物馆中大量的合影类照片,可以结合图像识别技术和知识图谱技术,快速检索照片中每个人物的背景信息,并将受关注的人物信息快速标注以展示给观众,同时延伸展示该人物的其他信息,这必将创新公共教育方式。

利用机器学习、神经网络等技术进行历史文物的深度挖掘和分析,结合诸如 Stable Diffusion 这样的人工智能绘图应用,配合人工干预手段,可以复现历史场景,还原历史文物的原始状态和属性。这开创了新的文物复原展览方式,为观众带来了更加生动的学习体验(Stability AI, 2023)。

五、文物大数据发展带来的挑战

大数据、人工智能应用给观众提供了更丰富、更灵活的游览体验,给博物馆事业发展带来了新助力,但也必将引发新的挑战。

首先是数据安全问题。在大数据时代,数据安全问题已经成为受到社会各界广泛关注的议题。个人隐私的保护在大数据时代面临着许多挑战,用户数据的收集、存储、分享和利用中都可能涉及信息泄露(刘雅辉、张铁赢、靳小龙等,2015)。大数据应用为观众制定个性化的游览内容和线路,必然涉及大量用户隐私数据的采集和分析,因此保护用户的个人隐私数据势在必行,这是一个备受关注的问题。为加强用户的隐私数据保护,可以采用多种技术手段,如数据加密、分散式架构或存储、匿名技术、隐私保护数据挖掘、差分隐私保护、数据访问控制等(刘雅辉、张铁赢、靳小龙等,2015)。考虑到隐私数据保护的复杂性,国家应在法律法规层面加快立法,并对现有法律法规进行修订,以法律手段确保用户隐私数据的安全。

其次是文物数据版权问题。文物数据的交互流通是文物活化利用的必然趋势,因此应重视文物数据的版权问题。国家层面应考虑制定完善的数据交易标准和规范,同时相关行业也应利用文物元数据的标准化来明确文物数据的版权归属。此外,技术标准如区块链技术标准等,也应明确引用为行业标准,并应用于文物数据的交易中(赵丰、周围,2017)。除了文物版权,还应关注文物形象所衍生出的 IP(intellectual property) 版权。近年来,许多知名博物馆成功打造了馆

藏文物相关 IP,如甘肃省博物馆打造的“神马来了”IP 系列就在网络上引起了热议(杨虞波罗,2022)。区块链等技术可以完善 IP 版权授权和归属认定问题,确保 IP 版权交易更加透明,这也是 IP 版权交易发展的拓展方向(陈维超,2018)。

最后,人工智能,核心在人。随着大数据应用的发展,博物馆行业对具备数据处理能力、模型优化能力的计算机学科人才的需求越来越大。针对博物馆领域的大数据应用,笔者认为模型优化的关键在于数据收集、数据预处理以及模型结果评判函数设计这三个阶段。虽然模型接受训练后可能具备处理多项任务的能力,但如果缺乏相关领域专业知识的人工辅助和评定,输出内容的质量也无法保证。有学者指出,盲目增大语言模型的训练数据并不能带来良好的效果,语言模型生成的结果仍可能出现虚假、有害或无意义的情况;这些学者利用人工评价所设计的奖励模型来迭代训练语言模型,显著提高了模型的性能(Ouyang, Wu & Jiang et al., 2022)。因此,笔者认为开展技能培训课程,培养具备文博专业知识的人工智能研究人员,使其具备数据处理、模型结果评判、损失函数或奖励函数优化等能力,对博物馆行业人工智能的发展具有重要意义。

六、文博领域大数据发展展望

保护历史文物是传承中华优秀传统文化的必然要求。党的十八大以来,党中央高度重视文物保护利用和文化遗产保护传承。习近平总书记强调:“要系统梳理传统文化资源,让收藏在故宫里的文物、陈列在广阔大地上的遗产、书写在古籍里的文字都活起来。”

《“十四五”文物保护和科技创新规划》明确提出,应建设国家文物资源大数据库,加强文物资源大数据应用,加强文物数字化保护,推动文物保护利用。本文认为,依托资源大数据库的建立,可以促使全国各博物馆、纪念馆、图书馆等收藏机构的文物数据得以真正流通;同时,通过预留功能性拓展接口,为将来与外部系统、平台的联通做好准备,实现全国各收藏机构文物的互动,为利用大数据挖掘分析和人工智能训练提供数据支持。

随着各行各业的数据积累和跨行业数据孤岛的打通,笔者期待大数据结合人工智能在文博行业、文旅行业实现通用式的大模型应用。这一应用可以根据游客的画像制定个性化旅游路线,包括预订机票、高铁交通、安排住宿、选择景点

线路、推荐餐饮、步行导航和购物索引等,从单个场景的“智能讲解”进化到覆盖整个旅游领域的“智能旅行助手”。政府部门也可以分析大量终端用户的反馈数据,实现对文旅行业中交通、场馆、餐饮、住宿、购物等环节更加精准的服务和管理。

参考文献

- 陈红京,2019,《博物馆学概论》,北京:高等教育出版社。
- 陈维超,2018,《基于区块链的IP版权授权与运营机制研究》,《出版科学》第5期。
- 邓慧超,2018,《试论博物馆如何利用大数据让文物活起来》,doi: 10. 26914/c. cnkihy. 2018. 009246。
- 富治平,2007,《江姐托孤遗书首度公开,筷子磨成竹签写成》,http://news. cctv. com/society/20071115/100222. shtml。
- 葛勇文,2021,《革命文物知识图谱构建研究》,河北大学硕士学位论文。
- 龚青,2011,《陈列展览策划与博物馆建设》,《东南文化》第4期。
- 国家文物局,2021a,《备案信息检索》,http://nb. ncha. gov. cn。
- 国家文物局,2001b,《博物馆藏品信息指标体系规范(试行)》,https://www. docin. com/p-1784261221. html。
- 国务院办公厅,2021,《国务院办公厅关于印发“十四五”文物保护和科技创新规划的通知》,http://www. gov. cn/zhengce/content/2021-11/08/content_5649764. htm。
- 国务院第一次全国可移动文物普查领导小组办公室、国家文物局,2017a,《第一次全国可移动文物普查数据公报》,http://www. ncha. gov. cn/art/2017/4/7/art_1984_139587. html。
- 国务院第一次全国可移动文物普查领导小组办公室、国家文物局,2017b,《第一次全国可移动文物普查工作报告》,http://www. ncha. gov. cn/art/2017/4/7/art_1984_139379. html。
- 红村网,2016,《杨虎城出国纪念锦旗》,https://bbs. ihcun. com/Searchdata/info/id/2878。
- 华龙网,2021,《七封江姐家书》,http://cq. cqnews. net/html/2021-10/24/content_51771184. html。
- 黄明玉、冯英、张俊娥等,2017,《文物分类体系(征求意见稿)》,https://www. lib. pku. edu. cn/portal/cn/news/0000001494。
- 李文昌,2008,《发展中的中国数字化博物馆》,《国际博物馆》(中文版)增刊第1期。
- 练洁、李娉、赵星宇,2021,《革命文物元数据标准研究》,《中国博物馆》第3期。
- 刘峤、李杨、段宏等,2016,《知识图谱构建技术综述》,《计算机研究与发展》第3期。
- 刘雅辉、张铁赢、靳小龙等,2015,《大数据时代的个人隐私保护》,《计算机研究与发展》第

1 期。

刘毅,2019,《文物学概论》,北京:高等教育出版社。

毛芳,2014,《浅谈可移动文物的数字化保护》,中国文物保护技术协会第八次学术年会论文。

王宏钧,2001,《中国博物馆学基础》,上海:上海古籍出版社。

杨虞波罗,2022,《“一见你就想笑”,丑萌“马踏飞燕”玩偶为啥能火出圈》,[http://ent. people. com. cn/n1/2022/0704/c1012-32465661. html](http://ent.people.com.cn/n1/2022/0704/c1012-32465661.html)。

张当中,1997,《汉字识别技术综述》,《语言文字应用》第2期。

赵丰、周围,2017,《基于区块链技术保护数字版权问题探析》,《科技与法律》第1期。

朱强、肖珑、张俊娥等,2017,《文物元数据体系(征求意见稿)》,[https://www. lib. pku. edu. cn/portal/cn/news/0000001494](https://www.lib.pku.edu.cn/portal/cn/news/0000001494)。

Baca, M. , P. Harpring & J. Ward et al. 2022, “Introduction to Metadata: Crosswalk.” [https://www. getty. edu/research/publications/electronic_publications/intrometadata/crosswalks. html](https://www.getty.edu/research/publications/electronic_publications/intrometadata/crosswalks.html).

Getty 2017, “Art & Architecture Thesaurus.” [https://www. getty. edu/research/tools/vocabularies/aat/index. html](https://www.getty.edu/research/tools/vocabularies/aat/index.html).

ICOM 2022, “Museum Definition.” [https://icom. museum/en/resources/standards-guidelines/museum-definition](https://icom.museum/en/resources/standards-guidelines/museum-definition).

Ouyang, L. , J. Wu & X. Jiang et al. 2022, “Training Language Models to Follow Instructions with Human Feedback.” ArXiv Preprint ArXiv:2203.02155.

Stability AI 2023, “Stability AI.” [https://stability. ai](https://stability.ai).

编委会主任：高岩

编委会副主任：夏桂华 赵玉新

吕鹏（中国社会科学院）

编委：尹航 冯仕政 冯全普

（按姓氏笔画排序）

吕鹏（中南大学） 吕冬诗

朱齐丹 汝鹏 苏竣

李正风 来有为 肖黎明

邱泽奇 何晓斌 宋士吉

陈云松 陈华珊 郑莉

孟小峰 孟天广 赵万里

赵延东 胡安宁 袁岳

黄萃 梁玉成 董波

曾志刚 蔡成涛 璩静

青年编委：丁奎元 王磊 叶瀚璋

（按姓氏笔画排序）

邢麟舟 向维 刘灿辉

刘松吟 刘春成 刘晓波

安博 许馨月 孙宇凡

李子信 李天朗 李晓天

吴雨晴 何丽 邹冠男

张咏雪 张承蒙 陈茁

陈典涵 林子皓 周雪健

周骥腾 郑李 胡万亨

茹文俊 贺久恒 贾雨心

郭媛媛 黄可 梁轩

曾晨

编辑团队

主编：郑莉

编辑部主任：吴肃然

编辑部成员：林召霞 王立秋

李昕茹 李天朗

岳凤

主管单位：中华人民共和国

工业和信息化部

主办单位：哈尔滨工程大学

出版单位：哈尔滨工程大学

出版社

地址：哈尔滨市南岗区

南通大街 145 号

国际标准连续出版物号：

ISSN 2097-2091

国内统一连续出版物号：

CN 23-1615/C

印刷单位：哈尔滨理想印刷有限公司

创刊年份：2022 年

出版日期：2023 年 11 月 10 日

发行单位：哈尔滨市邮局

订阅处：全国各地邮电局

邮发代号：14-375

发行范围：公开发行

定价：45.00 元

投稿指南

本刊面向海内外学者征稿，欢迎社会科学及交叉学科的专家学者惠赐稿件。请在来稿首页写明文章标题、作者简介（姓名、工作单位全称、联系电话、详细通信地址、电邮地址等）。文稿需完整，包括标题（中英文）、作者姓名、作者单位、摘要（300 字左右）、关键词（3—5 个）、正文、参考文献等。所投稿件如受基金资助，请在标题上加脚注说明，包括项目全称和项目批准号。来稿请以中文撰写。

稿件采用他人成说的，须在文中以括注方式说明出处，并在篇末列出参考文献；作者自己的注释均作为当页脚注。中外文参考文献分开列出，中文文献在前，外文文献在后，并按音序排列。中文文献参照中文社会学权威期刊格式，外文文献参照 APA 格式。来稿中的图表要清晰，符合出版质量要求，必要时可单独提供图表压缩包文件。

稿件格式请参考杂志官网（<http://www.jis.ac.cn>）“下载中心”中的稿件模板。

投稿方式：请登录杂志官网投稿系统（<http://www.jis.ac.cn>）进行投稿。

编辑部联系方式

地址：黑龙江省哈尔滨市南岗区南通大街 145 号哈尔滨工程大学主楼北楼 N301 室，《智能社会研究》编辑部

邮编：150001

电话：0451-82588881

E-mail: mailto:jis@163.com

著作权使用说明

本刊已许可中国知网等网络知识服务平台以数字化方式复制、汇编、发行、信息网络传播本刊全文。本刊支付的稿酬已包含网络知识服务平台的著作权使用费，所有署名作者向本刊提交文章发表之行为视为同意上述声明。如有异议，请在投稿时说明，本刊将按作者说明处理。