

深耕数据蓝海 赋能数智转型：

专业出版社高质量数据集建设探析

孙伯阳

在生成式人工智能(AIGC)技术快速迭代,文化数据要素市场化落地的双重背景下,高质量数据集已然成为文化产业转型升级的核心生产资料。作为垂直领域知识生产与权威内容传播的核心主体,专业出版社积淀的海量优质内容资源,是适配人工智能(AI)模型训练、智能知识服务的核心优质数据源,其开展高质量数据集建设,既是响应国家战略的必然举措,也是突破行业发展瓶颈、培育全新增长动力的关键路径。

数据集建设的时代必要性

2022年,《中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见》发布,明确将文化数据列为重点培育的数据要素。后续国家数据局启动高质量数据集建设专项试点,将文化、教育领域纳入优先试点范围,为出版行业数据资源活化利用指明了方向。专业出版社收藏的古籍善本、学术专著、行业规范、权威工具书等资源,是我国专业知识与文化数据的核心载体,将这些静态存量资源转化为标准化、可流通、可复用的动态数据要素,是出版行业落实国家数据战略、助力文化强国建设的核心抓手。

当前出版行业数字化转型已进入数智化深水区,传统的内容电子化扫描、简单格式转换等初级模式,已无法适配AI时代的产业需求。开展高质量数据集建设,能够系统性梳理、重构、规范存量内容资源,推动出版业从传统内容售卖模式,向知识服务、数据赋能的多元化业态转型。同时,数据集建设为出版产业开辟了全新盈利赛道。长期以来,专业出版社的优质内容仅依托纸质图书、电子读物、单机数据库等传统形式变现,资源复用率低、盈利模式单一,大量优质专业数据资产长期闲置。在数据要素市场化改革背景下,标准化、合规化的出版数据集可作为独立数据产品,为科研院校、AI企业、行业机构提供数据授权、定制化加工、专业知识服务等增值服务,有效打破传统出版业的盈利边界,激活垂直内容的核心商业价值。此外,专业出

版社深耕行业多年,参与行业标准建设,内容质量审核的经验丰富,牵头参与数据集建设,能够补齐文化数据行业标准短板,规范市场数据质量与交易秩序,推动文化数据市场规范化发展。

专业出版社的核心建设优势

相较于互联网平台、科技企业,专业出版社在高质量数据集建设中具备不可复制的天然优势,核心竞争力集中体现在内容、编审、生态与合规四大维度。

优质权威的垂直内容是出版社的核心底气,这也完美契合了高质量数据集“权威、准确、规范”的核心要求,是垂直领域AI训练的最优数据源。尤其是古籍类出版社珍藏的孤本、善本、民国文献等稀缺资源,经过长期专业校勘整理,形成了独一无二的低价位内容储备。而成熟的专业编审能力是数据集质量的核心保障。高质量数据集并非简单的文本堆砌,需要精细化加工与严格审核把控品质。出版社拥有学科背景扎实、行业经验丰富的编辑团队,能够深度参与数据标引、结构化重构、内容合规审核等工作。同时,能够依托完善的内容动态更新机制,持续校正数据疏漏、迭代内容,保障数据集的精准度与时效性。在生态资源层面,出版社深耕行业多年,在数据集建设方面具备强大的资源整合能力,能够牵头参与行业数据标准制定,吸引科技企业、科研机构参与数据开发与应用,构建良性产业生态。此外,出版社拥有完善的版权管理体系和稳定的作者合作渠道,能够精准界定版权归属,高效完成数字化使用授权,规范数据流通规则,有效规避数据合规风险,这也是其核心竞争优势之一。

行业建设的五大现实挑战

尽管优势突出,但目前专业出版社高质量数据集建设仍处于探索阶段,面临法律产权、版权授权、市场交易、技术能力、人才储备五大核心困境,制约了数据价值的有效释放。

法律产权界定模糊是首要难题。出版数据集创作涉及作者、出版社、数据加工人

员、技术开发方等多元主体,各方权利边界、收益分配方式目前还缺乏明确的法律界定。其次,版权授权与维权体系也存在明显短板。现有著作权法规难以适配AI训练、数据市场化流通等新型场景,大量老旧出版资源的原始授权协议未涵盖数字化深加工、智能应用等权限。同时,行业缺乏灵活的分级授权、收益分成机制,无统一的版权公共服务平台。而数据易复制、传播隐蔽的特性,导致新型数据侵权行为难、赔偿低、维权成本高。同时,市场交易机制缺失令数据价值难以落地。行业目前也暂无统一的数据评估与定价标准,确权登记、交易仲裁等配套机制不完善,跨行业数据流通壁垒未破除,数据集应用场景单一,难以形成“建设—交易—迭代—增值”的良性循环。此外,数据集建设需要大数据、自然语言处理、智能质检等前沿技术支撑,但出版社人才队伍以内容编辑为主,具备研发能力的技术团队不足,难以保障数据集标准化、高质量建设。整个行业也缺乏“出版+技术+法律+运营”的复合型人才。

构建全链条高质量发展体系

摆脱行业发展困境,需要国家、行业、出版社多方协同,从法律、版权、市场、标准、技术、人才六个维度系统发力,构建合规化、标准化、智能化、市场化的数据集建设体系。

首先,完善法律制度体系,筑牢合规根基。应加快适配AI时代的版权法规修订,明确出版数据数字化加工、AI训练等场景的合规边界,细化作者、出版社、加工方的权益划分规则与收益分配机制。同时,完善数据侵权判定、取证与赔偿标准,为数据集建设与流通提供法律支撑。其次,搭建行业公共平台,优化版权授权体系。由行业协会牵头成立数据集管理体系,搭建全国统一的出版版权数字化服务平台,实现授权、签约、结算、溯源全流程线上办理,解决授权效率低、作者失联等问题。同时,推行分级分类授权模式,针对科研公益、商业训练、定制服务等不同场景,制定差异化授权与收费标准,适配市场化多元需求。再

次,健全市场交易机制,激活数据资产价值。依托国家数据基础设施,搭建专业化文化出版可信数据交易平台,完善质量评级、价值评估、交易仲裁等全链条功能。行业统一数据定价标准,结合数据规模、稀缺性、应用价值等指标构建科学定价体系,破除跨领域数据流通壁垒,持续拓展数据应用场景,形成完整的产业闭环。同时,推进行业标准共建,实现数据规范化发展。构建政府引导、行业牵头、产学研协同的工作机制,对标国家及行业指南,统一出版数据格式、标引规则、质量检测、存储安全等细分标准。全面完成存量资源标准化改造,严格把控增量数据质量,打破数据孤岛,实现行业数据共建共享。在技术层面,强化智能技术赋能,升级数据加工能力。出版社需要加大技术投入,引入大数据、人工智能、语义分析等前沿技术,搭建智能化数据加工体系。建立全流程动态质检机制,运用加密存储、溯源存证等技术手段,全方位保障数据集质量与数据安全。最后,优化人才培养体系,补齐人才短板。构建内部转型培育、外部高端引进、校企联合培养的三位一体人才机制,对在岗编辑、技术人员开展交叉专业培训,引进数据技术、合规运营高端人才,完善激励机制。深化校企合作,定向培育复合型行业人才,为数据集长效建设提供人才支撑。

总之,高质量数据集建设是专业出版社拥抱数字经济、实现数智化转型的战略机遇,也是出版行业赋能文化强国、数据强国建设的重要抓手。依托权威的内容资源、专业的编审能力、深厚的行业生态等,专业出版社具备打造优质文化数据资产的核心能力。未来,随着相关制度体系不断完善、行业标准持续统一、技术人才短板逐步补齐,专业出版社将持续盘活优质文化数据资源,推动数据集建设向规模化、智能化、产业化升级,让传统出版资源在数字时代释放全新价值,为文化产业高质量发展注入持久数据动能。

(作者单位:国家图书馆出版社)

文旅融合背景下民族排舞的传承与创新发展

李奕霖

民族排舞融合了我 国各民族传统舞蹈韵律与标准化排舞形式,具有较强的观赏性,且群众参与门槛较低。作为优质的文体活动载体,近年来,民族排舞通过采风创作、景区展演、群众赛事等多元形式融入各地的文旅消费场景,成为诸多地方拉动文旅消费、传播当地民族文化的特色抓手。结合近年来民族排舞在不同类型的文旅场景中的落地实践案例,探索民族排舞的有效传承方式与贴合市场需求的创新发展路径,有助于更好地盘活我国各民族优秀文化资源,丰富地方文旅特色内容。

融入元素:深挖民族舞蹈内核,搭建在地化排舞内容资源库。民族排舞创作可围绕各地的民族文化资源开展定向采风,深度挖掘适合用排舞形式呈现的舞蹈、音乐素材,避免直接生搬硬套通用排舞的创作套路。比如,国家体育总局体操运动管理中心组织排舞采风团前往云南大理,走访当地多位非遗代表性传承人,梳理云南各民族传统舞蹈的動作与文化内涵,创编出《大理三月好风光》《霞光》《马樱花》等原创民族排舞。这些作品既保留了白族霸王鞭“蹓、蹓、蹓”的核心动作,阿昌族霞光“蹓、跳、蹓、扭、转”的标志性动作,也完整还原了每个动作对应的民俗生活意象,同步编制的中英文对照统一舞谱也被纳入中国特色排舞套路库,是民族排舞实现更广泛传播的代表性创新成果。

各地可以联合群众艺术馆、非遗代表

性传承人、排舞领域专业人士组建采风专班,每年选取当地世居民族的代表性舞蹈开展采风,同步做好民族舞蹈资源的数字化记录工作,为当地非遗资源完成系统的数字存档。创编完成的民族排舞还可以录制成标准化的教学视频,除了面向公众开放免费下载外,还可以将这些内容投放至本地的融媒体中心、社区公共公告屏、学校体育教研系统中,进一步拓宽这些作品的受众触达渠道。

强IP:依托专业赛事赋能,打通文旅融合消费闭环链路。借助民族排舞赛事自带的流量优势,可以联动本地文旅资源完成引流转化,打破赛事和文旅场景之间的壁垒,将单次赛事的短期热度转化为长效消费驱动力。比如,广西举办第四届排舞大赛期间,同步设置了潮玩市集,集结广西本土优质品牌参展,展售本地特色农产品、民族文创周边,并且,将该赛事的总决赛举办地放在了广西南宁的核心商圈,观赛市民参与积极性很高,实现了“以赛聚人,以玩促融”的活动初衷,有效带动了商圈人气上涨。

各地承办各级民族排舞赛事时,也可以同步策划赛旅融合配套活动,将赛事周边景区、特色村寨、网红餐饮商户纳入专属消费地图,游客凭参赛证、观赛票即可享受专属折扣;还可发起线上“民族排舞打卡挑战赛”,鼓励参赛者、游客加上专属话题发表相关视频,赢取景区年票、本地文创大礼包等,进一步扩大大赛传播覆盖面,带动

“吃住行游购娱”全链条消费增长。

育人才:夯实基层推广力量,构建全覆盖传播推广体系。要让民族排舞实现常态化普及,只靠优质内容与赛事支撑还不够,需要搭建分层级的民族排舞推广人才体系,保障其常态化落地普及。各地可参考广西排舞社会指导员的成熟培训模式,每年面向本地社区工作者、学校体育教师、广场舞骨干开展专项培训,培训内容兼顾排舞技术规范、本地民族文化常识、基层活动组织方法,还可邀请当地的非遗代表性传承人现场授课,讲解动作背后的民俗故事,让指导员不仅会跳,还能讲清其中的文化内涵。指导员还可以根据不同群体调整动作难度,比如,给老年人适配节奏更慢、动作幅度更小的版本,给青少年编排更富活动感的版本,形成“培训—推广—普及”的完整链路,覆盖全年龄段参与群体。

各地可以每年开展优秀推广指导员评选,给表现突出的个人颁发荣誉证书、文旅消费券等奖励,充分激发其推广积极性,并定期组织跨区域指导员交流活动,分享实操经验、碰撞创作思路。

造场景:创新景区沉浸式业态,打造常态化文旅体验名片。民族排舞可以结合本土特色资源,在核心景区设立常态化的民族排舞展演点位,安排本地排舞队伍定时展演,同步配套游客体验环节,免费教授民族特色排舞的基础动作,从而有效拉长游