

中信出版社数据资源管理研究

苏 曼

北京印刷学院出版学院，北京

收稿日期：2025年5月20日；录用日期：2025年6月17日；发布日期：2025年6月27日

摘 要

在数字化浪潮的推动下，数据资源管理已成为专业图书出版企业提升核心竞争力的关键要素。本文以中信出版社为研究对象，系统探讨其在数据资源管理构建中的战略规划、技术应用与组织创新。通过分析出版行业的数据管理现状及挑战，结合中信出版社的实践经验，本文揭示了其通过AI技术赋能、数据中台搭建等举措实现数据资产价值最大化的路径。研究进一步提出，出版企业需强化非结构化数据处理能力、构建开放的数据生态合作网络，并注重“技术 + 内容”复合型人才培养，以应对未来数字化竞争的复杂需求。本文不仅为出版行业提供了可复制的数据管理方法论，也为其他传统行业的数字化转型提供了重要参考。

关键词

数据资源管理，中信出版社

Research on Data Resource Management of CITIC Press

Man Su

School of Publishing, Beijing Institute of Graphic Communication, Beijing

Received: May 20th, 2025; accepted: Jun. 17th, 2025; published: Jun. 27th, 2025

Abstract

Driven by the digital wave, data resource management has become a key element for professional book publishing enterprises to enhance their core competitiveness. This article takes CITIC Press as the research object and systematically explores its strategic planning, technological application, and organizational innovation in the construction of a data resource management. By analyzing the current situation and challenges of data management in the publishing industry and combining CITIC Press's practical experience, this article reveals the path it has taken to maximize the value of

data assets through measures such as AI technology empowerment, data middle platform construction. The research further proposes that publishing enterprises need to strengthen their ability to handle unstructured data, build an open data ecosystem cooperation network, and focus on cultivating “technology + content” compound talents to meet the complex demands of future digital competition. This article not only provides a replicable data management methodology for the publishing industry but also offers an important reference for the digital transformation of other traditional industries.

Keywords

Data Resource Management, CITIC Press

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

1.1. 研究背景与意义

随着《关于构建更加完善的要素市场化配置体制机制的意见》将数据列为新型生产要素，出版行业的数字化转型进入加速期。据《2022 年中国数字阅读市场研究报告》显示[1]，中国数字阅读用户规模已达 5.3 亿，市场规模突破 500 亿元，数据驱动的精准运营成为行业竞争焦点。然而，传统出版企业在数据采集、分析及应用环节仍面临技术滞后、资源分散等痛点。以中信出版社为代表的头部企业，通过构建系统化的数据资源管理体系，实现了从内容生产到市场推广的全链路优化，其经验对行业具有重要借鉴意义。

1.2. 研究框架与方法

本文采用案例分析法与文献研究法，结合中信出版社年报、行业政策文件及学术研究成果，从战略、技术、组织三个维度解构其数据管理体系。通过对比行业现状，提炼其创新模式的核心价值，并提出面向未来的优化建议。

2. 专业图书出版行业数据资源管理现状与挑战

2.1. 行业数据管理现状

国家新闻出版署《出版业“十四五”发展规划》明确提出“建设行业级数据共享平台”[2]，极大地推动数据资产化进程。2023 年《出版业数据分类与安全规范》发布，为数据管理提供了标准化框架。在此基础之上，专业图书出版企业的数据资源类型与规模呈现出分层的趋势，头部企业如中信出版社、人民文学出版社已部署 AI 算法、云计算等技术，实现数据实时分析与决策支持，年新增数据量达 PB 级，中小出版社则仍依赖 Excel 等工具进行数据统计，缺乏系统化数据管理平台，年新增数据量普遍停留在 TB 级。其中，结构化数据包括图书销售记录(如当当网、京东平台实时销量)、库存信息(印刷批次、仓储分布)、财务数据(成本核算、利润率)等，占比约 60%；非结构化数据涵盖读者评论(豆瓣书评、社交媒体互动)、作者手稿(创作笔记、修改记录)、市场调研报告(用户访谈录音、视频素材)等，占比 40%，但利用率不足 30%。

2.2. 核心挑战与瓶颈

专业图书出版社数据孤岛问题较为突出，在出版社内部，编辑、营销、财务部门使用独立系统，数据格式不兼容(如编辑系统采用 XML 格式，销售系统依赖 CSV 文件)，跨部门协作效率低下；在出版社外部，出版社与作者、印刷厂、电商平台之间缺乏统一数据接口，导致市场反馈滞后。随之而来的是一系列的隐私风险问题，数字化时代，数字化内容容易被爬虫抓取，引发盗版问题(如某学术图书电子版在暗网流通量超 10 万次)，且读者评论情感分析、作者创作意图识别等场景缺乏成熟技术支撑，人工处理成本高昂，非结构化数据价值待挖掘。

3. 中信出版社数据资源管理体系构建路径

3.1. 战略规划：从“内容驱动”到“数据驱动”

中信出版社于 2018 年提出“智慧出版 2030”战略，以数据为核心重构出版全价值链，明确数据资源管理的三大目标：业务智能化、资产数字化、生态协同化，具体分为三个阶段实施。

3.1.1. 基础设施层(20182020)：夯实数据底座

中信出版社部署混合云架构，投资 1.5 亿元搭建“私有云 + 公有云”混合存储体系。私有云用于存储核心业务数据(如作者合同、版权信息)，公有云(AWS、阿里云)承载高并发访问需求(如电商销售数据、社交媒体流量)。至 2020 年底，数据存储容量达 5 PB，支持日均处理 10 万条交易记录。

同时，中信出版社还大力推出数据中台建设，构建“夸父数据中台”[3]，集成数据采集、清洗、存储、分析四大模块。中台采用微服务架构，支持快速扩展功能模块。例如，2020 年上线“实时销售看板”，可同步京东[4]、当当等 12 个平台销售数据，延迟低于 1 分钟。通过历史数据资产化，中信出版社对建社 30 年来的 20 万种图书资源进行数字化扫描，利用 OCR (光学字符识别)技术提取文本，构建结构化数据库。例如，《激荡三十年》等经典图书被转化为可检索的电子档案，支持 IP 二次开发。

3.1.2. 应用层(20212022)：场景化赋能业务

中信出版社引入 AI 驱动的选题决策系统，开发“选题智脑”模型(如表 1)，输入参数包括历史销量(权重 30%)、作者影响力(权重 25%)、社交媒体热度(权重 20%)、竞品分析(权重 15%)、政策趋势(权重 10%)。模型输出选题通过率预测值，准确率达 85%。2021 年通过该系统筛选的《碳中和革命》等 10 个选题，首印量均超 5 万册。

Table 1. Cases of AI technology application by CITIC Press
表 1. 中信出版社运用 AI 技术案例

案例	痛点	解决方案	成效
AI 选题决策系统	传统选题会依赖编辑经验，通过率不足 30%，且首印量预测偏差常达 40%以上。	开发“选题智脑”系统，集成 5 大类数据：历史销售数据(权重 30%)、作者影响力指数(权重 25%)、社交媒体热度(权重 20%)	2022 年通过率提升至 65%，首印量预测误差控制在 15% 以内。《元宇宙大爆炸》通过系统评估后首印 8 万册，3 个月内加印 3 次。
	印刷订单分散在 6 个基地，设备利用率不足 60%，交货周期长达 45 天。	部署 IoT 传感器采集设备状态数据、开发基于强化学习的调度算法	设备利用率提升至 85%，交货周期压缩至 30 天，2023 年节省成本 1200 万元。

同时设立动态库存管理系统、智能营销引擎，中信出版社基于 LSTM (长短期记忆网络)算法预测销量，结合实时销售数据动态调整加印策略。2022 年库存周转率从 3.2 次提升至 4.5 次，滞销图书比例下

降 18%；整合读者行为数据(阅读时长、点击热点)与外部环境数据(节假日、热点事件)，生成个性化营销方案。例如，2022 年“双十一”期间，系统自动为《底层逻辑》设计“满减 + 直播抽奖”组合策略，单日销量突破 3 万册。

3.1.3. 生态层(2023 至今)：构建开放数据生态

中信出版社积极构建“出版 + 数据服务”生态圈，推出“谷知谷知”知识付费平台，与得到 APP、喜马拉雅合作，将图书内容拆解为音频课、短视频等形态。平台集成学习行为分析功能，例如记录用户章节重复收听率，反哺纸质书再版优化。2023 年该平台营收占比达总收入的 12%。中信出版社建立跨界数据联盟，与抖音电商共建“出版短视频”数据闭环，通过用户观看时长、评论情感分析优化直播选品。2023 年“中信童书专场”直播销售额达 1200 万元，转化率较传统渠道提升 40%。

在行业标准输出方面，中信出版社主导制定《出版业数据分类与安全规范》，推动行业数据接口标准化。2023 年与商务印书馆、人民文学出版社等 10 家单位签订数据共享协议，实现选题查重、版权预警等协同功能；联合高校建立“出版大数据实验室”，探索 AIGC (生成式人工智能)在内容创作中的应用。

3.2. 技术应用：AI 与大数据深度赋能

3.2.1. 数据采集与整合：全链路自动化

中信出版社在数据采集与整合方面大力推进全链路自动化举措。在内部系统集成上，中信出版社通过 API 网关打通 ERP、CRM、SCM 系统，实现数据自动抽取。例如，编辑提交选题时，系统自动调用 CRM 中的作者信用评分、SCM 中的印刷成本数据，生成综合评估报告。

针对外部生态合作，中信出版社构建电商数据接口与社交媒体爬虫，与京东图书签订独家数据协议，获取用户“浏览加购购买”全路径数据，用于构建“购买意愿指数”；使用 Scrapy 框架抓取微博、小红书等平台话题标签，结合 NLP 技术提取关键词。如 2022 年，中信出版社通过分析“元宇宙”讨论热度，提前 3 个月策划《元宇宙大爆炸》并成为畅销书。

3.2.2. 数据分析与决策：智能化工具矩阵

在数据分析层面，中信出版社构建了“夸父 AI”平台，“夸父”平台的核心功能主要为素材生产、文本创作与润色、营销推广等。智能封面设计基于 GAN(生成对抗网络)生成多版设计方案，编辑可通过投票系统选择最优版本。2022 年设计周期从 14 天缩短至 3 天，封面点击率提升 25%。内容质量审核，利用 NLP 技术检测逻辑矛盾、敏感词与版权风险。例如，《全球通胀之谜》书稿中“美联储政策矛盾”段落被系统标记，经人工核查后修正。读者画像系统细分 8 大类 32 小类读者群体(如“Z 世代科技爱好者”“高净值投资人群”)，支持精准广告投放。2023 年定向推送的转化率较广撒网模式提高 35%。

在数据决策上，中信出版社推出了实时决策看板功能，管理层可通过移动端查看“销售热力图”、“库存预警指数”、“选题通过率趋势”等关键指标。例如，2022 年通过热力图发现长三角地区少儿图书需求激增，紧急加印《小狗钱钱》5 万册并迅速售罄。

3.3. 组织创新：敏捷架构与中台支撑

3.3.1. “举手制”工作室：激发微观活力

中信出版社于 2023 年获评“全国出版业数据化转型标杆企业”，承接国家新闻出版署“数据中台建设指南”课题。《出版人》杂志案例研究显示，中信出版社数据应用成熟度领先行业平均水平 35 年，这与其特别制度息息相关。

在编辑层面，公司实行“举手制” + “事业部制”。公司内部有能力、有想法的编辑可以申请独立成立工作室/子部门，工作室各自承担策划选题和收稿的工作，每周由主编、策划编辑、产品经理等组成

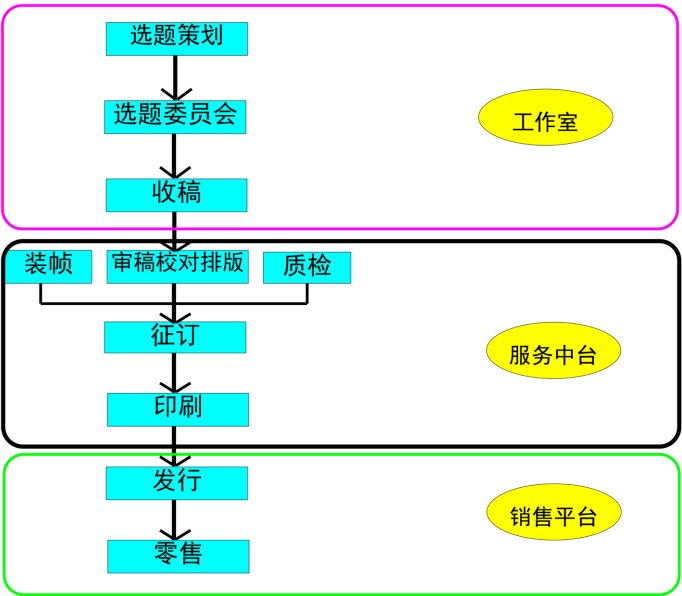
的选题论证会对所有选题进行筛选，后续工作室将持续跟进参与封面设计、单品营销等环节[5]。

中信出版社内，编辑可自主申请成立工作室，成员规模 515 人，实行“赛马机制”。2023 年全社共有 23 个工作室，贡献 70% 的畅销书。各工作室基于自身在不同专业出版领域积累的资源以及对市场动态的敏锐判断，可以自主确立选题方向。并且工作室将市场销售业绩和读者反馈作为关键绩效指标，促使编辑团队主动贴近市场，全力打造精品图书，这充分激发了自主创新活力。

同时，中信出版社采用扁平化管理结构，将工作室或子部门编辑团队规模一般控制在 9 至 20 人，这一规模便于团队负责人对选题策划、编辑加工、排版设计到印刷发行的出版全流程进行全面把控，及时调整策略解决问题。而且，经验丰富的核心编辑可高效地把专业知识、行业经验和出版趋势认知传递给新成员，加速人才培养，提升团队整体业务能力。其绩效考核与“三率”挂钩，选题通过率(30%)、首月销量(40%)、读者评分(30%)。例如，“财经图书工作室”因《价值投资实战手册》(首月销量 10 万册，豆瓣评分 8.9)获得年度奖金池的 25%。

3.3.2. 服务中台与销售中台：标准化支持与规模效应

中信出版社还将出版流程中具有规模效应的部分整合为服务中台和销售中台(如图 1)。这个服务中台涵盖质量管控、印制服务、IP 运营等职能，从前期市场调研数据支持，到中期质量把控，再到后期印制服务保障，可以助力工作室专注内容创作与营销策划，从而在经管、社科、少儿、文学等多品类实现高效、高质量出版，满足多样读者需求，为各工作室提供标准化、专业化支持。质量管控中台集成 AI 审校、印前质检功能。2022 年检测出 1.2 万处内容错误，人工复核效率提升 50%。印制服务中台通过物联网连接全国 6 大印刷基地，实现“一键下单自动排期进度追踪”。IP 运营中台管理 100 + 图书 IP 的衍生开发，如《人类简史》改编为纪录片、桌游，2023 年 IP 衍生收入占比达 18%。编辑提交选题后，服务中台自动分配资源，质量中台评估内容风险，印制中台核算成本，IP 中台规划衍生路径。例如，《元宇宙大爆炸》选题通过后，IP 中台同步启动了有声书、虚拟展览开发[6]。



Data sources: Prospectus of CITIC Press, Development Research Center of GF Securities.
数据来源：中信出版招股说明书，广发证券发展研究中心。

Figure 1. Organizational structure of CITIC Press
图 1. 中信出版社组织架构

销售平台承担市场分析、营销策划、渠道征订、销售推广、数据建模等工作。在选题策划环节,产品经理需介入选题筛选流程;印刷阶段,工作室将预先就拟出版图书的基础信息与销售平台进行前置性沟通,产品经理与销售经理基于此类信息完成潜在客户群体画像分析,并向目标客户开展订购意向征询工作,最终结合征订量统计数据及市场反馈信息,确定图书定价策略、首印数量及销售推广方案;销售环节中,公司通过自主研发的数据决策系统动态调控图书加印量与加印频率,该系统搭载的机器学习算法以企业线上/线下直销渠道及核心代销渠道的实时销售数据为输入变量。

4. 基于中信出版社成功经验的行业优化建议

4.1. 非结构化数据的深度挖掘与价值转化

当前,专业图书出版企业中非结构化数据占比高达 40%,但利用率不足 30%,其潜在价值亟待释放。未来需通过技术创新与场景适配,实现从“数据沉睡”到“价值激活”的跨越。

在未来,出版社应注意自然语言处理(NLP)的精细化应用,如读者评论情感分析采用 BERT、GPT4 等预训练模型,对豆瓣书评、微博话题进行情感极性分类(正面、中性、负面),识别高频关键词。例如,针对《经济学原理》的读者评论,模型可提取“案例不足”“公式晦涩”等反馈,指导再版时增加案例库与注释模块。在进行作者创作意图挖掘时可以通过 LDA (潜在狄利克雷分布)主题模型分析作者手稿与创作笔记,提取核心思想与知识关联。

中信出版社在应用场景拓展方面具有独特的贡献,这对出版社的未来发展提供了建议。在未来,出版社可注重智能荐书系统升级,基于知识图谱与用户行为数据,实现“千人千面”推荐。例如,对科研用户推荐“理论 + 实践”组合(如《机器学习基础》+《Kaggle 竞赛实战》),对职场新人推荐“技能 + 案例”组合(如《高效沟通》+《腾讯产品方法论》)。出版社可积极引入强化学习(RL)优化推荐策略,根据用户点击率、购买转化率动态调整权重,推荐准确率提升 15%~20%。

4.2. 新兴技术融合与场景创新

出版业需跳出传统技术框架,探索区块链、元宇宙、量子计算等前沿技术的融合应用,构建“技术内容体验”三位一体的创新生态。

实现区块链技术的场景化落地,需要出版社做到版权确权与交易透明化,将图书版权信息(作者署名、授权范围、分成比例)上链存证,实现不可篡改的确权记录[7]。例如我们可以借鉴中信出版社与蚂蚁链合作,为《人类简史》电子版生成唯一数字指纹,任何未经授权的传播均可溯源追责[8]。出版社可搭建去中心化版权交易平台,支持作者、出版社、IP 运营方直接交易,减少中间环节成本。通过智能合约自动执行版税分成,纠纷处理周期从 3 个月缩短至 7 天。同时,应注意大力进行 NFT 数字藏品的多元化开发,可提供限量版数字藏品,将经典图书封面、作者手稿扫描件转化为 NFT,赋予收藏价值。如《原则》作者达利欧亲笔签名版 NFT,在阿里拍卖平台以 12 万元成交。出版社还可以加入互动式阅读体验,读者购买 NFT 后可解锁专属内容(如未公开章节、作者访谈视频),增强用户粘性。

元宇宙场景的沉浸式探索也是新时代的必由之路。通过虚拟图书展销会,在 Decentraland 等元宇宙平台搭建虚拟展厅,用户通过数字分身参会,实时采集动线数据(停留时长、交互热点),优化展位布局。在元宇宙应用探索的道路上,中信出版社积极进取,成果颇丰。2024 年,中信出版集团与中信书店携手全球知名科普品牌 DK 于北京举办了“致·好奇心——DK50 周年自然典藏百科展”,这一展览系 DK 独家授权的全球首展,含金量极高。华星创业与英国 DK 公司、中信出版集团合作,创新性地将元宇宙沉浸式体验融入其中。在展览现场,设置了 5 大声光电互动体验区域,运用数字科技打造出 VR 恐龙世界和沉浸式宇宙星空,为观众带来了震撼的感官体验,华星创业全资子公司洸泠社提供了本次展览中元宇

宙 VR“光影万象”展区的全部内容和技术方案,融合 VR、AR 及全息技术,构建起多维度、多互动且富有启发性的科普内容[9]。

步入 2025 年,在腾讯云 AI 引擎的有力支持下,中信出版社成功打造出国内首个出版元宇宙“CITIC Verse”。用户只需佩戴 VR 设备,便能仿若身临其境般“走进”《清明上河图》的场景之中,还能与历史学者虚拟人展开对话。此外,用户购买实体书后可自动解锁专属数字资产,这种虚实结合的创新模式极具吸引力,“CITIC Verse”上线首月便成功吸引 Z 世代用户超 80 万,在出版行业元宇宙应用领域树立了新的标杆[10]。

4.3. 复合型人才培养与组织能力升级

中信出版社的成功经验启示出版社,数据资源管理体系的可持续性依赖于“技术 + 内容 + 商业”跨界人才的储备。需从教育体系、企业内部机制等方面突破。

教育体系改革方面,通过高校培养和职业培训可以为出版社培养高素质复合型人才[11]。可通过开设“出版数据科学”联合学位、建立“出版科技实验室”、职业培训体系创新等方法,或推出“数据驱动出版师”认证,要求候选人掌握 SQL 查询、Tableau 可视化、基础机器学习算法,并通过“选题策划数据模拟”考核;与 Coursera、得到 APP 合作开发在线课程,覆盖数据采集、分析、合规等全链路技能。

企业内部要进行机制优化,参考中信出版社“璞玉计划”,“璞玉计划”是中信出版集团的青年人才发展项目,致力于选拔全球一流院校的优秀毕业生,通过入职培训、岗上教练带教、专项培训、项目历练以及轮岗进行职业赋能,内容涉及公司企业文化、职业素养、业务全貌等多个维度,为公司未来业务发展引进高素质、高水平、高学历人才。广大图书出版社可开展“数据先锋”人才计划,设立专项奖金激励员工考取 CDMP(数据管理专业人士)、AWS 数据分析认证等资质,2023 年中信出版社已有 15%编辑通过认证。出版社根据实际情况,可实施“轮岗实践”,要求数据工程师参与选题论证会,编辑团队参与数据清洗项目,促进跨职能理解。同时引入外部专家,聘请数据科学家担任“出版顾问”,指导技术落地。参考中信出版社经验,出版社可与咨询公司合作开展“数据成熟度评估”,每年发布《出版数据能力白皮书》,明确改进方向。

5. 总结

数字化转型浪潮下,出版业正经历深刻变革,数据资源管理成为核心驱动力。中信出版社的实践为行业数字化转型提供了示范路径。

其“智慧出版 2030”战略构建了全方位变革框架。战略层面以数据重塑出版全流程;技术层面 AI 深度赋能选题、库存等关键环节;组织层面通过“举手制”工作室和服务中台实现高效协同。这一模式凸显了技术业务深度融合解决痛点、三位一体创新构建良性循环、开放生态实现价值共创的深刻启示。

未来出版业需重点突破挖掘非结构化数据价值、融合区块链、元宇宙等新技术、培育复合型人才等方面。通过持续创新,出版业将实现转型升级,为文化传播注入新动能。

基金项目

出版智库平台建设,项目号:KYCPT202514。

参考文献

- [1] 艾瑞咨询. 2022 年中国数字阅读市场研究报告[R]. 2022.
- [2] 本刊编辑部. 从“十四五”到“十五五”: 出版业的转型与升级[J]. 科技与出版, 2025(1): 5.
- [3] 中信出版集团. 中信出版 2020 年度报告[R]. 北京: 中信出版社, 2021: 15-18.

- [4] 李文彤. 大数据背景下中信出版社内容资源的数字化开发[J]. 西部广播电视, 2019(6): 43-44.
- [5] TMT 研究-爱好者. 深度|中信出版: 组织架构、内容与渠道的成长性探讨(广发传媒互联网旷实团队)[Z/OL]. 雪球网. <https://xueqiu.com/9231373161/137391938>, 2019-12-15.
- [6] 元宇宙产业联盟. 2023 出版业元宇宙应用案例集[R]. 2023: 33-35.
- [7] 国家新闻出版署. 出版业区块链技术应用白皮书(2022 版) [R]. 北京: 人民出版社, 2022: 35-38.
- [8] 蚂蚁链研究院. 区块链在数字出版领域的应用实践——以《人类简史》为例[R]. 杭州: 蚂蚁集团, 2021.
- [9] 银柿财经. 携手中信出版带来 VR “史前世界”华星创业元宇宙业务加速商业化落地[EB/OL]. <http://cn.beijingleather.com.cn/jiaodian/2024/0801/279442.html>, 2024-08-01.
- [10] 雪球网. 中信出版在当下 AI 的机遇与挑战[EB/OL]. <https://xueqiu.com/2303270949/323525021>, 2025-02-15.
- [11] 范军, 肖璐. 出版融合背景下传统出版企业人才队伍建设的困境与对策[J]. 中国出版, 2016(23): 19-23.