

基于图表资源的科技论文多元化传播探索实践*

宫在芹 朱拴成 武英刚

煤炭科学研究总院出版传媒集团,100013,北京

摘要 图表是科技论文的重要组成部分,图表资源更是科技期刊提供知识服务的重要数据源。科技期刊图表资源具有复杂性、存量巨大且持续增长、可重复使用、价值链未形成等特征。在分析国内外典型图表数据库的基础上,介绍煤炭科学研究总院出版传媒集团图表库的建设背景、建设目的、实施路径,以及建设过程中的难点和解决对策。通过图表资源的标准化加工、结构化存储等激发了论文传播的内生动力,有利于科研成果的多元化传播。

关键词 科技论文;图表资源;多元化传播

Communication exploration and practice of diversified scientific papers based on chart resources//GONG Zaiqin, ZHU Shuancheng, WU Yinggang

Abstract Figures and tables are important parts of scientific papers, and are important data source for scientific journals to provide knowledge service. The resources of figures and tables have the characteristics of complexity, large stock, continuous growth, reusable and immature value chains. Based on the analysis of typical figures and tables database in China and abroad, we introduce the database construction background, purpose, implementation path, as well as the difficulties and solutions in the construction process of Publishing and Media Group, China Coal Research Institute. The standardized processing and structured storage of resources for figures and tables stimulate the endogenous power of paper dissemination, which is conducive to the diversified dissemination of scientific research results.

Keywords scientific paper; chart resource; diversified dissemination

Authors' address Publishing and Media Group, China Coal Research Institute, 100013, Beijing, China

DOI:10.16811/j.cnki.1001-4314.2020.06.009

随着现代信息技术和出版文化的发展,人们获取信息的基本手段已开始从文字阅读逐渐转向文字和图像并重的阅读方式,图像、图形、视频等渐渐成为超越文字符号的阅读语言^[1]。图表是科技论文的重要组成部分,它可以形象、直观地表达科学思想和技术知识,并具有活跃和美化版面的功能,进而提高读者的阅读兴趣和阅读效率^[2]。科技期刊历来重视对图表的编校、审读。文献[1,3-6]分别对图表加工应坚持的原则、图表审读的注意事项、表格编校的设计优化、图

表版权与再利用等问题进行了研究。

与期刊出版前编辑人员投入大量精力编校图表不同,近年来,凭借技术和资金优势,以及科学的组织管理体系等,国内外大型出版商、数据库运营商等更加注重对出版后图表数据的资源化利用。作为我国第一个学术类图片知识库产品,中国知网学术图片库通过采用智能挖掘技术从CNKI总库中提取图表,通过规范整理,为用户提供图表的检索、对比分析等知识发现功能。依托刊群优势,中国地理资源期刊网的图表库集合了《地球信息科学学报》《地理科学》等11种期刊2012年至今的6万余幅图表,图表按所在期刊、发表年度进行分类。用户可在集群网站上检索并免费阅读目标图表、图表所在论文及文中的其他图表。Geofacets地质数据库隶属于Elsevier,其收录了地震剖面、地层学、井数据等方面的超过100万幅图表^[7]。由于石油和天然气市场不稳定以及学术研究预算减少,地球科学领域研究人员在发现和格式化地质数据中面临重大挑战。通过Geofacets地质数据库,研究人员可以快速检索到可靠的地理学出版物发布的地理参照图表,为其决策提供可靠的数据支持,提高生产率。隶属于细胞出版社的Cell Picture Show图片资源库由日本滨松光子学株式会社提供支持,资源库分几十个专题展示各种细胞与分子生物学领域的精美显微照片,读者可在注册后直接提交科研中积累的精美照片^[8]。

网络数字时代,如何充分挖掘优质图表资源的潜在价值,对其进行关联并动态重组,增强研究成果被发现的可能,进而提高论文的显示度是科技期刊健康可持续发展不能忽视的问题。本文以煤炭科学研究总院(简称煤科总院)出版传媒集团所属科技期刊为例,介绍期刊在图表资源知识标引、深度加工、结构化存储,并依托中国煤炭行业知识服务平台重组再利用,进而推动科技论文多元化传播的探索实践,希望能为其他科技期刊充分挖掘图表资源传播优势提供参考。

1 科技期刊图表资源特征

1.1 复杂性

科技期刊图表资源的复杂性,表现在图表的种类、格式复杂多样,质量参差不齐,关联关系复杂。图片种类有曲线图、形态图、谱线图、分析图、系统图等;格式

* 中国煤炭科工集团有限公司科技创新创业资金专项面上项目(2018-2-MS030);中国煤炭科工集团有限公司科技创新创业资金专项面上项目(2019-RC002)

上,有的期刊采用的是tif格式,有的则是jpg格式;绘图时,有的期刊为了保证印刷时的清晰度,像素设置得较高,如《煤炭学报》要求实物图和试验过程中拍摄的照片分辨率均不低于600 dpi,有的期刊采用的图表则清晰度较差。在对排版数据的标引上,大多数期刊没有对图片文件标引方式做出明确的规定,而排版人员为了提高效率,通常按照自己习惯的方式标引。以煤科总院出版传媒集团所属期刊为例,有的期刊以第一作者姓拼音首字母+图片序号进行标引,如第一作者为王国法的论文的第一幅图片被命名为wgf1;有的以稿号+图片序号进行标引,如2020-0121-1;有的期刊因有多位排版人员,同期图片出现了不同的标引方式。这种无序混乱的标引方式加上图表自身的复杂性,给其资源化应用带来了巨大的挑战。

1.2 存量巨大持续增长

据《2018—2019中国数字出版产业年度报告》统计,互联网出版商同方知网的《中国知识资源总库》目前已累计整合出版国内外文献2亿多篇,其中图片5000万张,且以每年150余万张的幅度增长^[9]。不仅互联网期刊出版商累积了大量图表资源,编辑部存储的过刊图表资源也数量可观。以《煤炭学报》为例,其每期的图表量已达到500余幅,自1964年创刊以来的图表量已累计近10万幅,且仍在以每年近万幅的速度增长。与文献资源可以通过期刊官网、互联网出版商进行分类有序出版不同,由于缺少专业可靠的图表数据深加工、融合发布管理平台和科学有效的内容管理技术与方法,使得大部分期刊对挖掘手中越来越多的图表资源缺乏动力。

1.3 可重复使用

网络时代的“信息爆炸”与“信息匮乏”增加了科技论文被发现和阅读的难度,研究显示50%甚至更多已经发表的文献从未被阅读过^[10]。作为科技论文的重要组成部分,在遵守版权保护的前提下,对图表进行合理引用或再利用,不仅可以节约科研成本,还可以增加原始图表所在科技论文被发现和阅读的概率,促进成果传播。生物医学领域的论文,特别是综述论文引用或再利用已发表论文图表的情况较多^[6]。图表不仅可以在学术文献中再次使用,还可以用于非出版领域,作为独立的作品存在^[1]。以煤炭行业的地质图为例,随着技术的进步和地质学的发展,地质图的绘制越来越精确,表现的内容也更加丰富。其他图表如,“2005—2016年我国煤矿事故死亡人数对比”“2008—2018年我国职业病新发病例与煤矿事故死亡人数统计”“智能化煤矿复杂巨系统逻辑关联图”等,这些图表均具有重要的参考价值,在征得编辑部/出版商或作

者许可,注明作者和出处的情况下可以再利用。

1.4 价值链未形成

在利润驱动下,大型出版商通常对图表资源有清晰的利用途径及产品包装形式,并可以根据市场需求,快速进行产品迭代,以满足用户需求。反观国内大多数期刊,受制于有限的人力与紧张的办刊经费,其对已经拥有版权的图表资源的服务价值不清晰、处理环节不明确、整合处理能力薄弱,导致科技期刊图表资源化发展方向不明确。探索图表资源价值链成为科技期刊图表库发展的必然,构建清晰、多元化的图表价值体系,成为科技期刊图表资源开发的方向。

2 煤科总院出版传媒集团所属期刊图表库建设实践

2.1 建设背景

2.1.1 自身发展需要 中国煤炭行业知识服务平台是由中国煤炭学会学术期刊工作委员会与煤科总院出版传媒集团联合主办的煤炭行业唯一科技期刊集群门户网站,平台于2016年上线,至今先后经历多次改版,在媒体融合、服务科研等方面做了许多探索,积累了宝贵的经验^[11]。平台已经建成的论文库、专题库、专家库、视频库、问答库通过多角度、立体化地呈现学术成果及专家风采,有效提升了煤炭行业科技期刊的数字化进程及影响力^[12]。在近5年的资源库建设中,平台累积了数量庞大的图表资源和较为丰富的数据处理经验。为了标准化、产品化图表资源,使用户在平台上可以多维度地获取知识链接,进一步提升行业期刊影响力,2019年底平台启动图表库建设。

2.1.2 行业经验驱动 新时代,文化出版产业呈现出精品化、多元化、创意化、规模化等发展态势^[13],这对科技期刊实现融合发展,从学术出版向知识服务转型提供了巨大的推动作用。如爱思唯尔、美国化学会等均通过精选数据内容资源并进行结构化、知识化深加工,打造了多种服务形态和产品形式。中国知网学术图片库、中国地理资源期刊网的图表库、中华医学期刊全文数据库、Geofacets、Cell Picture Show等图表资源库的建设运行表明,与传统出版的线性传递相比,经过标引、重组的图表资源可以更加精确而充分地满足读者的个性化需求,促进论文的多元化传播。

2.2 建设目的

2.2.1 挖掘内部潜力,激活传播内生动力 煤科总院出版传媒集团共有期刊20种,其中直管期刊5种,分别为《煤炭学报》《煤炭科学技术》《洁净煤技术》《国际煤炭科学技术学报(英文)》(JCST)和《智能矿山》。20种期刊中11种为核心期刊,9种为非核心期刊;其

中《煤炭学报》、JCST 为 EI 收录,且在 2019 年入选“中国科技期刊卓越行动计划”;《煤炭学报》《煤炭科学技术》《煤田地质与勘探》为 CSCD 收录期刊。

近年来,在建设世界一流科技期刊等文件及政策的支持鼓励下,出版集团的高影响力期刊持续创新,通过组织优秀专题、与高端学术论坛合作、积极拥抱新技术等措施,影响力越来越大,论文被引频次和影响力等评价指标逐年升高。与之相比,虽然影响力与知名度不高的期刊也在努力尝试通过各种新媒体平台技术推广优秀论文,但阅读率、转发率、评论率等指标均不理想,面临着学术影响力提升难的窘境。作为研究成果的重要载体,做好图表资源的利用,不仅可以多元化传播高影响力期刊的优秀成果,还可以提升影响力较低期刊论文的显示度,激发论文传播的内生动力。

2.2.2 立足自身优势 满足用户需求 煤科总院出版传媒集团所属期刊的报道范围涵盖煤炭开采、运输、转化、利用、环保等全产业链,内容倾向于工程应用,读者群主要为生产一线的技术人员。近年来,随着煤矿规模化、智能化水平提高以及煤矿装备更新换代加快,生产企业对解决生产实际问题的科研成果有越来越大的需求,如井下锚杆锚索怎样布置、巷道如何支护、瓦斯抽采钻孔漏气如何补漏等。许多生产中遇到的难题,有时通过科研成果中一副图表就能为技术人员提供很大的启发。但由于出版形式及检索机制等的限制,传统的按照主题、期刊、篇名、摘要等方式进行的知识检索已经很难帮助生产一线科研人员快速便捷地找到解决方案。通过将出版集团所属期刊的图表进行碎片化标引、重组、入库,并设置多维的检索机制,可以更好地满足用户生产需求。

2.3 建设路径

2.3.1 图表的标准化处理 高质量、标准化的图表是图表分析和应用的基础。中国煤炭行业知识服务平台图表库是一个开放、多维、动态的数据库。因为 20 种期刊分布在北京、上海、重庆、西安、常州、抚顺、唐山、杭州、淮北等地区,有合作多年的排版公司及相对成熟的编校流程,为了不给各刊额外增加工作负担,图表库的建设由出版集团新媒体事业部统一负责,各刊将存量图表数据及新出的图表发至新媒体事业部。在图表库建设中,出版集团制订了详细的图表处理规范,对图表的格式、元数据结构等做了规定。入库图表所使用的元数据结构由 11 项元素组成,涉及图表名称、作者、论文标题、学科、基金等。

2.3.2 与其他库的数据交互 随着技术的发展,单一的论文下载、孤立的图表展示已无法满足科研工作者的需要,因此图表库在建设之初就将与其他资源库的

交互作为重点工作。图表的标准化处理中“论文 id”描述保证了其与平台上指定论文的链接。经过标准化处理后的图表按照数据入库标准进入图表库,并通过匹配论文 id 自动与论文库实现数据交互。在后台,图表与论文的编辑功能在同一编辑词条中,对需要更改的图表可以方便地替换修改。在前台,图表与论文既独立存在,又相互关联。在论文库的论文详情页中,标准化处理后的图表在摘要下方折叠展示,丰富了论文信息;在图表库中,图表作为主体,通过学科细分,突出了图表的个体属性。图表库与问答库等的互联通过超链接的方式实现。

2.3.3 图表库外观设计 平台在构建图表数据库时,充分从用户的角度出发进行设计,认为图表库的主要功能是为科研人员提供研究环境并帮助其发现问题,而不仅仅是储存与检索。因此在图表库的主界面中,按照学科进行分类,弱化期刊、作者、基金项目等因素,引导用户将关注点聚焦到学科内容。在检索上,用户既可以利用图表库主界面顶部的学科汉语拼音首字母排序查找图表,也可以按照图表名称、学科、作者等进行检索。在检索结果页中,设置了期刊、作者、发表年、受资助基金等维度信息,供用户进一步筛选。图表全面完整信息的表达仍有赖于所在论文,所以图表详情页最终指向论文库中论文源,供用户免费下载阅读。在浏览效果上,对于单幅图表,考虑到在 PC 端及移动端的显示效果,对图表高度做了统一处理,保证用户在 PC 端浏览的大多数图表能显示在一屏中,不用上下滑动鼠标;在移动端,为了适配移动读图场景,默认是小图。需要浏览图表细节时,在单击图片后,设计了手势缩放功能。

2.4 建设难点及建议

2.4.1 文件的收集 期刊在生产过程中产生的各种出版数据、版式文件等电子文件是出版社最重要的数字资产,也是重要的战略储备,在数字技术时代具有潜在的使用价值^[14]。但许多期刊并没有建立科学规范的数据归档、存储规定,再加上人员流动、设备更换等原因,都出现了电子文件保存不全的问题,甚至很多期刊将原始图表等数据交给印刷商保管,而自己只存储正文 PDF 数据。在图表标准化处理中,tiff 格式的文件因为可以高质量地完成转换和输出而被作为首选文件格式。但在文件收集,出版传媒集团的很多期刊都无法提供创刊至今所有的 tiff 格式原始图片。面对这种情况,在制作时不得不采用从 PDF 正文中截取图片的方式来完成图表的标准化加工,这不仅增加了工作成本,还会导致图表质量的下降。因此建议期刊在生产过程中重视数字资产的管理,尤其是 tiff 格式图

表的存储管理。

2.4.2 图表的学科划分 在知识总量迅猛增加,学科划分日益精细化的今天,对每一幅图表所属学科进行精准描述,不仅能提高其被检索的概率,更有利于促进交叉学科的发展。出版传媒集团图表库在建设之初就确定了在一级页面按照学科进行图表分类展示的目标。在实施过程中,最初拟定了2种确定图表所属学科的方案:一是由专业编辑确定;二是邀请专家列出学科列表,由新媒体事业部对照学科列表对图表进行分类。以上2种方案在试行时出现了较多问题:一是专业编辑时间有限,不可能对上万幅图表进行学科确定;二是很多学科名称是专业编辑依据经验来划分,名称出现不统一等问题;三是虽然专家列出的学科很详细,但在落实到具体图表时,如果缺乏专业知识,将很难完成准确归类。为了提高图表库建设效率,同时统一标准,降低人为因素的干扰,最终选择依据每篇论文的中图分类号确定该论文中每幅图表的学科。实践证明,这种分类大大降低了对图表加工人员的专业要求,可供其他期刊在图表整理时参考。

2.4.3 图表源文件与索引文档的匹配 图表的名称、作者、所属论文题目等描述信息依靠索引文档来记录,每一张图表对应一条索引信息。索引文档依靠图表名称识别图表源文件,并将其导入图表库,所以图表源文件的统一规范命名至关重要。但正如本文1.1中提到,期刊对图表源文件的命名并不重视,各刊之间的图表命名规则也存在很大的差异,在图表标准化制作中,不仅要先按照每本刊不同的排版规则,在图表源文件中找到目标图表,还要将图表名称改为与正文一致,这为图表的标准化制作增加了很多的工作量。融合出版的显著特征是一次制作,多次使用。期刊在出版过程中,应尽量按照统一规范方式处理图表,虽然暂时会增加工作量,但随着数据量的逐渐增多,这种事前的功夫会对以后的数据利用起到事半功倍的效果。

3 图表库建设成效

中国煤炭行业知识服务平台图表库自2020年4月上线以来,已累计为出版集团期刊标准化加工图表17 000余幅,涉及600余个细分学科,与学科热点相关的“煤矿全生命周期绿色开采技术框架”“智能化采煤工作面评价中各指标权重”“煤与瓦斯突出危险精准辨识原理示意”等图表受到了广泛的关注。图表库的建设成效主要体现在以下2个方面:一是引起了各刊对图表资源规范化制作、存储的重视,各刊按照出版集团对图表标准化加工的要求,正逐步完善图表的加工制作标准,并有序存储每期新出版的图表文件;二是增

强了论文传播内生动力。以图片“新冠病毒水中胶体颗粒构造示意”“新冠病毒在城市水循环系统中关键环节、关键位置存在方式及潜在传播途径”为例,这2幅图均来自于非核心期刊《能源环境保护》2020年第3期的论文,因其反映了当下的研究热点且信息完整度高,在图表库中获得了很高的关注度,激发了论文传播的内生动力。

4 结束语

科技期刊图表库的建设将完善图表的数字化与数据化工作,实现期刊图表资源与文献数据库的共同发展,为科技论文的推广传播提供新的思路与方法,为科研人员的研究提供系统全面的资源信息。在数字化时代,科技期刊应充分挖掘利用图表资源优势,推进图表数据资源向更加全面、细化、深入的方向发展。

5 参考文献

- [1] 张小强,钟紫红,张秀峰. 科技期刊论文插图的版权与再利用[J]. 中国科技期刊研究, 2012, 23(3): 429
- [2] 陈浩元. 科技书刊标准化18讲[M]. 北京: 北京师范大学出版社, 1998: 117
- [3] 张福颖,倪东鸿. 科技论文中图表编辑加工的8类情形[J]. 编辑学报, 2019, 31(4): 391
- [4] 陈先军. 科技期刊论文的图表审读处理方法探讨[J]. 编辑学报, 2018, 30(3): 266
- [5] 王小辰,蔡斐. 材料类科技论文表格设计常见错误剖析[J]. 编辑学报, 2017, 29(1): 39
- [6] 叶维萍. 生物医学期刊论文中再利用国外期刊已发表图表的授权许可[J]. 中国科技期刊研究, 2014, 25(6): 778
- [7] 中国科学技术协会. 中国科技期刊发展蓝皮书(2018)[M]. 北京: 科学出版社, 2018: 23
- [8] 肖宏. 面向国际的科技期刊数字化出版: CNKI数据库“走出去”[Z]. 北京: 中国农业科学院, 2019
- [9] 张立,王飏,李广宇. 2018—2019中国数字出版产业年度报告[M]. 北京: 中国书籍出版社, 2019: 104
- [10] 初景利,盛怡瑾. 科技期刊发展的十大主要态势[J]. 中国科技期刊研究, 2018, 29(6): 533
- [11] 宫在芹,朱拴成. 煤炭行业科技期刊服务科研的创新实践[J]. 编辑学报, 2019, 31(6): 669
- [12] 朱拴成. 科技期刊集群化服务平台融合出版探索实践: 以中国煤炭期刊网为例[J]. 编辑学报, 2019, 31(2): 209
- [13] 中国科学技术协会. 中国科技期刊发展蓝皮书(2020)[M]. 北京: 科学出版社, 2020
- [14] 金东,游苏宁. 科技期刊排版中心应重视数字资产的管理[J]. 中国科技期刊研究, 2007, 18(4): 583

(2020-08-07收稿;2020-10-13修回)