

国家工程技术图书馆“十四五”发展思考

曾建勋 周杰 杨代庆 王星
(中国科学技术信息研究所, 北京 100038)

摘要: 简要概述国家工程技术图书馆发展现状, 分析科技文献资源的出版和利用环境变化、科技文献的战略安全保障与服务等一系列变革对国家工程技术图书馆的挑战和机遇。在此基础上, 提出国家工程技术图书馆需要着眼于国家创新发展战略需求、国家科技信息安全、国家科技信息保障利用提升、国家科技信息服务均衡发展, 从资源建设、知识组织、知识服务、科研能力等方面部署相关重点任务, 继而形成国家工程技术文献发现服务中心、科技论文信息监测服务中心、文献情报基础设施研发推广中心、国家科研档案收藏管理中心“四大服务中心”。

关键词: 国家工程技术图书馆; 科技文献保障; 知识组织; 知识服务

中图分类号: G251 **DOI:** 10.3772/j.issn.1673-2286.2021.05.004

引文格式: 曾建勋, 周杰, 杨代庆, 等. 国家工程技术图书馆“十四五”发展思考[J]. 数字图书馆论坛, 2021(5): 17-22.

科技文献是建设创新型国家不可或缺的重要支撑资源, 持续稳定地建设科技文献, 对于我国科技文献战略安全至关重要; 科技文献的生产、组织、传播与利用方式已经发生颠覆性变革, 数字资源成为科技文献出版与利用主流, 科技文献信息呈现出多来源、多模态、多结构特点。为此, 图书馆的资源建设、数据组织、服务模式等各环节都面临机遇与挑战, 国家工程技术图书馆沿着既定“一体两翼, 图情档一体化融合发展”路径, 制定“十四五”期间发展规划, 进一步全面提升科技文献信息保障与服务能力。

1 国家工程技术图书馆发展现状

“十三五”期间, 国家工程技术图书馆继续夯实印本文献保障, 重点开展“文献大元数据体系”和“开放学术资源体系”建设, 进一步加强《汉语主题词表》、名称规范文档等基础信息设施的建设, 启动科技项目库、科技奖项库等事实型数据库的建设工作, 基本形成公益服务与情报服务相互促进、协同发展的新局面。

1.1 构建大元数据体系, 实现多源异构资源关联组织

截至2020年底, 馆藏及采购中外文各类文献文摘总量达2亿条, 引文总量达3亿条, 全文总量已达1亿个, 并与科睿唯安、Elsevier、Springer、ProQuest等27家出版社/集成商达成元数据合作, 获取相应元数据, 覆盖了Web of Science和Scopus等重要数据库的元数据。

作为NSTL的骨干成员单位牵头开展了“国家外文开放学术资源”建设, 从开放资源元数据标准制定、遴选规范、数据采集流程方法、版权评估及风险等多个方面进行研究, 并对1万种期刊、5 600种会议录、8万份学位论文、1.4万册图书、5 000份报告等国外开放资源进行元数据及全文采集保存, 总量超过700万篇。开展“国家科技期刊开放平台”建设, 通过逐一与期刊社签署协议获取授权方式, 目前实现1 200多种中文科技期刊的开放获取。

2016年成立科技档案业务组以来, 制定档案数字化加工标准和规范, 开发档案整理数字化加工平台, 接受科技部重大专项办公室的委托, 编制《国家科技重大专项项目档案管理工作方案》。截至2020年底, 完成

国家科技计划项目档案加工整理6万多卷,近10万盒。

同时启动“全球科研项目数据库”与“国内科技奖项数据库”建设,目前项目库涉及400多个国内外重要科研基金项目,项目信息超过700万条;奖项库收录国家级、省部级、学协会奖项数据20万条。多源异构科技信息数据经过自主建设的“元数据集成管理系统”集成整合,加载入大元数据仓储,初步实现主题、机构、时间的交叉关联组织。

1.2 启动联合目录建设,构建分级服务的文献发现系统

在NSTL的支持下,作为牵头单位联合上海图书馆、中国科学院文献情报中心、CALIS管理中心以及中国医学科学院医学信息研究所/图书馆等机构在既有联合目录及馆藏数据的基础上,启动“国家外科技期刊联合目录”建设,支撑国家科技文献保障服务体系。

为满足用户个性化、精准化的文献资源发现与获取服务需求,国家工程技术图书馆建设的科技文献发现服务系统(以下简称“发现系统”)正式上线运行,发现系统采用了新的大数据平台架构,支持海量文献数据的集成检索与分析,实现了2亿篇文献元数据与3亿条引文数据的集成关联,提供文献篇级、品种级等多维度资源“一站式”发现与全文获取。发现系统实现了用户的“阶梯式”实名认证策略与分级信用管理,支持手机、身份证认证,也支持IP地址、ID账号登录访问,还支持个人用户、集团用户等多粒度用户控制。发现系统中嵌入自主知识产权的《汉语主题词表》,对海量文献进行标引,支持主题概念检索与推荐聚类。发现系统采取微服务架构,实现文献服务、数据服务、情报服务各服务单元的松耦合分布式模块化设计,既可独立运行,也便于扩展升级,支持面向第三方提供快速嵌入服务。

1.3 强化名称规范档建设,夯实科技论文引文库基础

基本形成以机构名称、会议名称、期刊名称等为主体的各类实体名称规范档体系,并支撑国家工程技术图书馆的文献资源及增值情报服务,提升了对实体识别的效率,实现了实体归一的规范控制。名称规范档建设中已经制定完成了一系列文献著录标准和数据归一规

范,形成了相对完善的元数据集和规范名称集。目前,机构规范档库建设中,已经规范国内一级机构4300余个,其下属的二级、三级等机构52万余条,覆盖中国SCI论文第一作者机构发文94%以上,覆盖中文核心期刊论文第一作者机构发文75%以上。

基于构建的名称规范档的控制,包含6000多种中文学术期刊,2020年论文量5000多万篇,引文2.8亿条数据研发的中文科学引文索引数据库(CSCI)正式推出上线并提供服务,是目前最完备的中文论文引文数据库。CSCI数据库不仅提供文献检索,更重要的是一款基于引用的评价工具。经过对论文主题、作者、作者机构、期刊、引文和基金等信息相互关联和规范控制,共同支撑知识评价。基于CSCI数据库,每年编制出版《中国高被引分析报告》,按论文从属学科类别统计,从作者、机构、图书会议等多个角度分学科进行高被引分析,全面展示各个学科领域的高被引情况。

1.4 推进《汉语主题词表》全学科建设,促进知识组织体系落地应用

“十三五”期间,国家工程技术图书馆编制完成《汉语主题词表(自然科学卷)》,并印刷出版。《汉语主题词表(自然科学卷)》覆盖自然科学12个学科专业,收录正式主题词(叙词)约6.5万条,非正式主题词(入口词)6.1万条,概念属性包括范畴、英文译称、同义词、上位词、下位词、族首词、注释等。在完成《汉语主题词表》(工程技术卷、自然科学卷)的基础上,2019年获得科技部“科技基础资源调查专项”支持,2020年正式启动《汉语主题词表》生物卷、医学卷、农业卷的编制。

2018年底,《汉语主题词表》网络服务系统正式面向公众开放试用,同时提供网络接口服务。《汉语主题词表》服务系统是基于《汉语主题词表》的术语服务系统,提供术语检索、文本分词、主题标引、学科分类、知识树构建等服务。《汉语主题词表》服务系统突破了同类系统只提供术语检索服务的局限,是我国第一个同时具备知识组织工具发布和应用服务功能的术语服务系统。随着《汉语主题词表》服务系统进一步推广使用,有助于提升我国专业领域文本信息处理的自动化和智能化水平。

1.5 开拓服务新模式, 文献服务与专题服务协同发展

以国家工程技术图书馆发现服务为依托, 开展公益文献服务和培训, 文献传递量逐年加速上升, 2020年文献传递量达80万篇, 实现了较好的社会效益。为助力解决我国科技文献服务中存在的“不平衡”现象, 加强面向高职院校、欠发达地区的宣传推广, 并着重向行业拓展延伸。面对用户个性化需求, 快速构建“日文文献专题库”“航空航天文献专题库”等专题文献数据库, 提供普惠文献服务。

“十三五”期间, 国家工程技术图书馆积极面向各类创新主体的情报信息需求, 利用资源优势 and 各类数据库, 开展专题情报数据分析服务, 情报服务的用户包括高校、政府、企业, 开展的业务涉及学术评估、信息监测、专题研究等方面, 在服务过程中不断完善自身的能力, 提升产品质量, 已经逐步形成情报服务产品体系, 服务内容和模式逐步向实现公益服务与情报服务并重的新模式转型。

2 国家工程技术图书馆面临挑战与机遇

面对百年未有之大变革, 图书文献领域的挑战与机遇并存, 一方面, 科技文献资源的出版和利用已经发生颠覆性变革, 科技文献资源的全新数字信息资源产业链正在形成, 数字出版、语义出版、社交网络正成为信息资源生产的主流, 基于大数据的数据密集科研正成为科学发现的新范式, 用户对多来源异构数据的有机整合、知识发现、关联计算、关系推理提出了更高的需求; 另一方面, 国际地缘政治风险增大, 国外科技文献的战略安全保障与服务产生新的不确定性, 对传统资源采集渠道、服务手段等一系列环节带来新的挑战。国家工程技术图书馆作为我国科技文献重要保障机构, 要继续加强国内外各文献信息机构的合作, 在合作创新中发展, 最终发挥好图书情报事业在国家创新驱动发展战略中的支撑保障作用。

2.1 大数据、云计算和人工智能带来颠覆性变革

大数据、云计算、人工智能等新兴技术的迅猛发

展给科技文献保障体系建设带来前所未有的技术挑战。大数据的资源化加速了科技文献信息数字化进程, 对图书馆的数据存储能力、计算能力、知识组织能力、数据分析能力提出严峻挑战, 并将极大地改变图书馆信息资源建设与服务的现有模式^[1]。云计算作为以数据为中心的新型网络计算方式, 将深刻地影响未来图书馆的互联网运作和服务模式, 改变图书馆信息获取、存储、检索和共享的方式与理念。当前人工智能已成为引领未来的战略性技术, 在深刻改变人类社会生活的同时, 也给图书情报行业带来革命性变化。人工智能在智能参考咨询、智能检索、语义数字图书馆、图像识别等领域的直接应用, 将推动科技文献信息保障和服务从“互联网+”进入“智能+”时代。总而言之, 各类新技术对图书馆的技术应用能力和知识服务能力带来前所未有的技术挑战, 必将颠覆图书馆传统业务模式, 推动图书馆转型, 催生图书馆新业态, 释放新动能。

2.2 微服务和共享经济形成开放融合新机制

微服务作为一种架构, 已经成为技术热点, 微服务将单一应用程序划分成一组小的服务, 服务间互相协调, 为用户提供最终价值。在下一代图书馆服务平台的构建中, 普遍采用了微服务架构, 数据不再以集中式的中央存储为中心, 而是根据实际情况结合不同数据存储方式满足业务需求, 基于微服务架构的下一代图书馆开放服务平台将是对图书馆平台的一次重新设计, 将图书馆馆员的智慧、软件人员的技术、系统服务人员的市场能力融合在一起, 改变以往的合作模式, 使各方共同谋划图书馆的未来^[2]。

共享经济正在深入发展到各个行业之中, 颠覆社会相关产业发展模式。近年来数字图书馆领域基于互联网平台, 也在掀起共享潮流。社会化标注将分类主题标引任务外包给网络上的大众; 开放数据推进信息采集和资源整合; 开源软件在强化各类软件、工具和平台的开发提升。通过众创、众包、众扶、众筹等共享经济新模式正在实现跨界融合、跨界组织和跨界服务^[3]。

数字图书馆更需要勇于面对开放融合潮流的挑战, 推进开放共享与融合, 改善共建共享机制, 提升信息服务效率, 实现新常态下文献信息资源建设与服务的新转型。

2.3 中美科技博弈带来的科技信息安全风险

随着我国对科技文献资源保障投入的持续增加,我国科技文献事业得到了长足发展,然而应清醒地认识到:数字时代到来给我国的科技文献安全带来新的风险。随着科技文献从印本向数字资源转变,文献资源订购的仅是“使用权”,而非“拥有权”,数字信息资源的国家拥有和保存问题更加严峻。通过互联网在线访问国外数字文献资源所留痕的日志信息,有可能暴露我国科研动态和研究方向^[4]。随着中美科技博弈的加剧,发达国家对我国科技文献信息限购日趋加强,我国对外文科技文献访问和使用将遭到重大阻碍。我国缺乏高质量的科技期刊、会议录等原生产品,外文文献信息分析工具与系统多数依赖国外,国家科技文献保障体系在保证印本资源订购的同时,对纯电子出版的数字资源、网络资源、开放获取(Open Access, OA)资源等进行完整采集和本地保存,并实现本土化平台访问和保障迫在眉睫。

当前在科技文献领域,文献资源建设中的“结构不合理”和“服务不平衡”矛盾仍然存在,在部分地区还非常突出,主要集中在:文献资源采集未进行严格评估,导致文献采集后利用率不高;文献资源重复建设现象仍然比较严重,在资源建设中存在跟风采集现象,如SCI数据库的高校订户达到326家,远远超过“双一流”高校数量。此外,文献资源的保障以学术论文为主,对于产业发展有重要作用的咨询报告、事实型数据库保障力度不够,对于资源采集主要依托于传统商业数据库、出版商渠道,对于新媒体等的学术资源关注度与采集不足。

文献资源服务中不平衡现象更加突出,“双一流”高校的文献服务得到较好保障,但对于广大欠发达地区、高职高专师生,文献获取渠道和数量都极为有限。此外,作为创新主体的广大企业科研人员往往难以及时获取到所需要的科技文献资源。

3 国家工程技术图书馆“十四五”发展思考

国家工程技术图书馆定位于保障工程技术领域的科技文献国家战略安全,采集、长期保存各类工程技术创新文献资源,是国家工程技术文献基础设施的重要构建和维护机构,主要着眼于国家创新发展战略需

求、国家科技信息安全、国家科技信息保障利用提升、国家科技信息服务均衡发展,以公益文献服务为基础,积极开拓知识服务,支撑国家创新发展战略。把科技信息保障作为科技自立自强的战略支撑,把科技信息高端交流平台作为国家战略科技力量的重要组成部分加以建设。“十四五”期间将从资源建设、知识组织、知识服务、科研能力等方面完成相关重点任务。

3.1 建立大元数据库集成体系,实现多源异构资源融合

全面实施科技文献元数据战略,最大限度构建国家元数据库,探索基于知识产权管理、市场化激励措施的元数据登记注册集成共享机制,大力发展元数据多来源获取渠道,加强对网络资源、开放资源、订购资源、科研原生资源等元数据及时发现、采集、规范和保存,实现图书馆、出版社、集成商等多源异构文献元数据的集成融合,编制“外文科技文献联合总目录”,建立中国科技信息资源的“大”元数据体系。按照元数据收割/导入、转换、校验、集成、查重、归一等环节设计元数据集成融合流程,制定元数据统一标准规范和描述模型,实行多来源元数据格式的映射登记,建立元数据集成融合系统,实现多源元数据在数据、信息和语义3个层面的集成整合,形成覆盖元数据资源采集获取、集成整合、信息服务等全流程的功能模型,并不断提高元数据采集完整性和更新及时性,形成国家元数据库的可持续发展机制^[1]。

3.2 推进资源采集渠道建设,构建立体化资源保障体系

科技成果产出呈现资源立体化特征,资源类型极大丰富,积极开展资源采集的多渠道建设,除继续强化商业文献订购渠道的稳定外,大力推进网络资源的采集获取、拓展交换、会员代购、文献互助等多种文献获取渠道。除传统科技文献保障建设期刊、会议、学位论文、科技报告等一次、二次文献外,加强对新媒体、科学数据、可视化图谱等新兴资源的关注,强化对富媒体学术资源、事实型数据资源、术语型组织资源、软件型工具资源等建设保障,形成立体化资源保障体系;在全面推进面向学术研究的资源建设基础上,开拓面向产业创新的资源建设,加强市场报告、产业分析、统

计手册等资源采集,形成学术信息资源与产业信息资源相融合、适应知识增长和创新需求变化的一体化创新资源体系,有序地向多元化载体和新型信息内容资源建设拓展。

3.3 推动国家知识库建设,强化文献资源本土保障自主可控

建立国家(机构)知识库建设,推动国家财政支持的学术论文成果存缴。加快实现国际重要科技文献数字资源本土长期保存,形成数字全文长期保存和原文获取服务于一体的数字资源长期保存体系与协同服务保障格局;强化“卡脖子”资源的监测,尽快梳理摸清可能被限制而尚未实现本土化保存的重点领域国外重要科技文献,有步骤地进行采集和本土化保存;提升数字资源长期保存的自我管理能力,实现我国原生数据资源、引进国外数字资源、网络开放资源等的长期保存和增量可持续更新,建立异地备份机制,建立可靠的公共服务触发管理机制,增强抵御信息风险的能力。

3.4 夯实知识组织体系,深化语义标引与知识计算型数据处理

继续更新维护《汉语主题词表》(工程技术卷、自然科学卷),在科技基础资源调查专项支持下,大规模开展《汉语主题词表》的生物卷、医学卷、农业卷建设,在“十四五”期间,完成理、工、农、医四大自然科学门类的《汉语主题词表》建设,并进一步夯实“汉语主题词表在线服务系统”的功能和应用落地。在《汉语主题词表》的支撑下,加强数据处理语义标引及知识计算工具建设和集成,采用人工智能、大数据等技术手段,加强数据知识化组织,对数据进行深入的标引和挖掘。通过术语表示的知识内容和概念,按照统一的语义框架,将语义关系与各类资源紧密关联,形成多维度、多层次的资源深度聚合,实现资源的语义化、智能化的检索和展示^[5],以满足科学研究、技术创新、绩效管理和资源建设评估等多层次、全方位的需要。

3.5 完善资源发现系统,发展情景敏感的协同调度服务

不断优化和完善科技文献资源发现系统,逐步将

网络服务系统、知识链接系统、词表服务系统以及开放资源集成服务平台等进行数据、服务和系统层面的深度整合,研发包含数十亿级科技文献信息的大数据发现服务系统,形成统一集中的科技文献资源聚合服务体系,提供从信息搜索到全文获取的一站式公益性服务,成为全国的工程技术科技文献资源发现中心。扩展科技信息资源的发现途径和渠道,建设基于开放链接的资源配置与调度系统,根据用户使用环境,实现不同类型数字资源和不同服务方式之间的开放动态链接,集成印本资源、OA资源、面向全国公益机构开通的电子资源、回溯资源等,提供包括本地获取、全文传递、馆际互借、OA下载、回溯下载、DOI链接、单篇文献订购等情景敏感的多途径资源获取方式^[6],成为全国的工程技术科技文献资源配置调度中心。

3.6 基于用户实名注册,发展多元化的服务推广模式

大力推进用户实名注册管理,构建包括用户实名注册、机构契约注册在内的统一的用户管理系统,针对不同用户群体特征,实现用户信誉管理和分层级信息服务措施,在IP与ID控制下提供端对端、点对点服务,拓展文献信息服务规模;采取多样化的推广模式,建立用户行为数据库,开展用户画像,支撑精准宣传推广;制订统一规范的用户实名注册管理规定、工作流程、岗位责任,保护个人隐私和遵守知识产权,面向全社会推进泛在、公平、均衡、无障碍地发现获取与开放共享服务。拓展支持第三方的数据服务需求,支持第三方机构对国家工程技术图书馆资源与服务进行二次开发和深度定制,形成高附加值资源,开发和推广嵌入式资源利用工具,培育形成下一代开放工程技术领域知识服务体系新格局。

3.7 基于关联计算,发展专题情报分析服务

通过人工智能技术与语义知识组织的优势互补和深度融合,突破从科技文献、科学数据等可信赖知识源中进行元数据集成、数字知识表示、语义组织关联和用户认知模型构建方法,形成统一集中的科技文献资源聚合管理服务体系。强化对现有文献资源的揭示力度,加强对非析出单篇型资源的揭示,探索采用目录、章节知识点、内容提要等方式的揭示方法,并筹划表格、图

片、公式、概念等的资源细粒度组织和加工。在文献的关联性描述与揭示、各类实体等规范文档识别与规范基础上,通过对文献知识点细粒度的加工关联,推进高附加值资源库建设。拓展、深化信息情报服务,针对重点领域、重要行业,前沿与热点、技术应用实施等方面进行持续监测,提供科技文献、知识专题服务平台,形成丰富的情报产品体系,服务于更多用户。

4 结语

国家工程技术图书馆“十四五”期间将遵循“一体两翼,图情档融合发展”原则,面向科研创新主体,推进“基于知识组织的资源发现系统”建设,提升覆盖全国的服务能力;针对科技管理决策主体,组建信息资源分析团队,提升信息监测和知识服务能力;针对学科建设和行业领域,完善《汉语主题词表》和机构名称规范档建设,加强《数字图书馆论坛》建设,规范“全国信息与文献标准化技术委员会”管理,提升图书情报学科的科研能力。继而形成“四大服务中心”,即国家工程技术文献发现服务中心、科技论文信息监测服务中心、文献情报基础设施研发推广中心、国家科研档案收藏管

理中心。为此,需要从思想理念、业务流程、经费投入、人才体系、业务合作、科研支撑等多个方面予以条件保障,需要争取加大数字资源投入,完善信息服务平台和资源评估;强化信息安全意识,拓展信息获取与合作渠道;推进数字业务流程再造,构建联合协作生态;夯实文献情报科研基础,强化国际国内业务交流合作。

参考文献

- [1] 曾建勋. 开放融合环境下NSTL资源建设的发展思考[J]. 大学图书馆学报, 2020, 38(6): 63-70.
- [2] 肖铮, 林俊伟. 用微服务构架下一代图书馆服务平台——以FOLIO为例[J]. 图书馆杂志, 2018, 37(11): 63-69.
- [3] 曾建勋. 借力“四众”开创资源共享新模式[J]. 数字图书馆论坛, 2016(1): 1.
- [4] 曾建勋. “十四五”期间我国科技情报事业的发展思考[J]. 情报理论与实践, 2021, 44(1): 1-7.
- [5] 彭以祺, 吴波尔, 沈仲祺. 国家科技图书文献中心“十三五”发展规划[J]. 数字图书馆论坛, 2016(11): 12-20.
- [6] 曾建勋. 基于发现系统的NSTL用户服务体系思考[J]. 情报杂志, 2020, 39(11): 134-138.

作者简介

曾建勋, 男, 1965年生, 信息资源中心主任, 研究馆员, 博士生导师, 研究方向: 知识组织与知识链接, E-mail: zeng@istic.ac.cn.

周杰, 男, 1964年生, 信息资源中心副主任, 研究馆员, 研究方向: 信息资源建设。

杨代庆, 男, 1975年生, 信息资源中心副主任, 高级工程师, 研究方向: 信息资源建设。

王星, 男, 1977年生, 高级工程师, 研究方向: 信息管理与信息系统。

The 14th Five-Year Plan for the Development of National Engineering and Technology Library

ZENG JianXun ZHOU Jie YANG DaiQing WANG Xing

(Institute of Scientific and Technical Information of China, Beijing 100038, China)

Abstract: This paper briefly summarizes the current development of National Engineering and Technology Library, analyzes the challenges and opportunities to the National Engineering and Technology Library caused by a series of changes in the publishing and utilization environment of scientific and technological literature resources, strategic security and service of scientific and technological literature. On this basis, it is proposed that the National Engineering and Technology Library should focus on the needs of national innovation and development strategy, national science and technology information security, national science and technology information guarantee and utilization improvement, and balanced development of national science and technology information service, and deploy relevant key tasks from the aspects of resource construction, knowledge organization, knowledge service, and scientific research capacity, then form the “Four Service Centers” of the national engineering technology document discovery service center, the scientific paper information monitoring service center, literature and information infrastructure research and development promotion center, and the national scientific research archives collection management center.

Keywords: National Engineering and Technology Library; Scientific and Technological Literature Guarantee; Knowledge Organization; Knowledge Service

(收稿日期: 2021-04-23)