

国家科技文献保障工作的形势与 NSTL“十四五”规划

彭以祺

(国家科技图书文献中心, 北京 100038)

摘要: 科技文献是重要的科研条件和战略资源, 本文在简要概述我国科技文献保障工作形势和现状的基础上, 分析当前面临的困难和挑战, 提出国家科技图书文献中心 (National Science and Technology Library, NSTL) “十四五”规划总体发展目标是科技文献信息资源保障能力显著增强, 科技文献信息服务水平和智能化知识服务能力显著提升。规划确定了NSTL未来发展的五大任务, 即建设安全的国家科技文献本土化保障基地, 全面实施科技文献元数据战略, 强化系统和网络基础设施建设, 完善国家科技文献信息普惠服务和情报服务网络, 探索科技文献智能知识服务关键技术。

关键词: 国家科技文献保障; NSTL; 信息服务; 知识服务

中图分类号: G251 **DOI:** 10.3772/j.issn.1673-2286.2021.05.001

引文格式: 彭以祺. 国家科技文献保障工作的形势与NSTL“十四五”规划[J]. 数字图书馆论坛, 2021 (5): 2-7.

科技文献是以文字、符号、图形、音视频等手段记录科技信息、成果、知识的载体, 是对人类通过科学劳动和技术研发而获得的科学知识 with 文明进行记录和总结, 从不便利状态变为可利用状态, 并广泛应用于科技创新活动中的一种重要基础资源。科技文献是科技创新重要的基础性条件之一, 它不仅是简单的借阅和参考, 更渗透在科技创新活动中的每个环节, 并在循环利用中不断产生新的价值。科技文献是重要的科研条件和战略资源, 是科技创新重要的基础保障。目前, 正在研究和编制的2021—2035年国家中长期科学和技术发展规划中, 已将科技文献与科学数据、科研试剂、实验动物一并纳入国家创新平台及基础设施建设的规划之中。2020年4月29日, 科技部、财政部等六部委联合发文《新形势下加强基础研究若干重点举措》, 明确提出要“构建完善的国家科技文献信息保障服务体系”^[1]。《中共中央关于制定国民经济和社会发展第十四个五年规划和二〇三五年远景目标的建议》也提出要“构建国家科研论文和科技信息高端交流平台”。在科学技术飞速发展的今天, 科技文献已成为保障国

家科研环境安全、满足科研人员需求变化、提高科技竞争力水平、建设创新型国家不可或缺的支撑条件。

1 我国科技文献保障工作的形势和现状

1.1 我国科技文献需求十分旺盛

中国科技人力资源排名世界第一, 发表国际科技论文数及被引用次数世界第二, 专利数量多年世界第一。每年毕业的大学生、研究生高达850万人, 2021年高校毕业生人数将首次突破900万。当前, 中央和地方各级政府高度重视科技工作, 科研经费持续增加, 科技活动旺盛, 2019年我国R&D经费投入总量为2 2143.6亿元, 居世界第二, 与美国差距逐步缩小。因此, 中国是世界上科技图书文献用户最多、需求最旺盛、需求量增长最快的国家, 是科技文献信息需求最大的市场。我国每天有几千万人次在不同平台上查询、点击、阅读、下载科技文献, 如NSTL网站平均每天点击近149万次, 中国科学院的网络版全文下载量在2020年达到6 937万篇,

国家图书馆2019年每天访问量大约有220万次,中国知网每天访问量500万次,万方数据主站每天访问量200万次以上,百度学术每天访问量超过700万次。

另外,科技文献的应用面不断拓宽,过去科技文献主要为科学研究、教育教学服务,但现在随着科技创新工作的深入开展,在申报国家重点研发项目和重点基金、开展研究工作和技术开发、高等教育与研究生培养、国家高新企业创新发展、科学普及、全民教育以及社会发展、经济建设、气候变化等各个方面都离不开科技文献信息的支撑。

1.2 科技文献呈爆炸式增长

伴随着科学技术的不断发展,各个学科领域的研究与实践日益增多,科技文献和数据呈现爆炸式增长。据统计,目前全球约有学术期刊6.5万种,其中约有2.3万种为STM同行评议期刊,这些期刊每年出版的论文约140万篇。目前我国每年出版中文科技期刊约5 000种,中文科技图书约16万种,学位论文约73万份。另外,数字资源发展迅猛,已成为科技文献资源的主要出版形态。总体而言,科技文献尤其是科技期刊已经实现印本与电子版本并存出版和电子版优先出版,并逐步走向纯电子出版。科研用户对当前数字信息环境的依赖性越来越强,电子期刊已成为科研教育机构用户使用的主要资源,电子科技专著也迅速成为科研教育机构用户使用的重要资源。众多信息服务机构积极推动将信息资源建设转向数字资源。数字文献资源的长期保存得到高度重视,出版界、图书馆界等跨界合作形成了多个多方合作的长期保存方案,包括芬兰国家图书馆和德国国家图书馆的国家保存体系,也包括市场化合作保存的Portico系统和CLOCKSS系统。不过,为规避仍然存在的风险,具有国家保障性质的图书馆并未主动减少印本资源的订购,旨在建立稳定可靠的国家战略信息资源。

1.3 科技文献服务平台更加专业化、知识化

最近十年,我国科技文献服务平台不断增加和完善,主要可以分为两类。一类是公益性的科技文献服务平台,包括国家图书馆、国家科技图书文献中心以及医药卫生、农业科学、机械、化工、冶金等专业文献服务平台。公益类科技文献服务平台具有资源量大、内容

丰富、服务多样、开放共享程度高、公益普惠等特点,服务范围广,用户众多,是我国科技文献服务的骨干力量。另一类是市场主导的商业类科技文献网站和平台,如中国知网、万方数据、维普、百度学术等,形成独特的文献资源、强大的技术实力和灵活的市场机制,在科技文献服务特别是中文文献领域占有一席之地。这些服务平台基本满足了不同层面科技工作的需要。

同时,我国科技文献服务平台的建设在专业化、知识化方面迈出坚实步伐,文献服务平台内容和功能不断丰富,包括检索获取、原文传递、计量分析、精准发现等,并加强对创新、管理、决策的支撑服务以及面向企业、产业的保障服务。另外,在数据密集型科学范式下,也在积极探索科技文献的智能化服务,提供个性化服务。图书馆成为巨大的原生数字内容的管理者、新数据集合的创造者和分析者。NSTL组织开展了下一代开放知识服务平台总体设计及关键技术研发,开展RDF语义映射动态配置和大规模自动转换引擎、科技论文摘要语步标注工具、领域术语抽取自动化流程与算法、语义智能检索与认知搜索工具、基于用户增强画像的个性化推荐工具等研发,推动下一代开放知识服务平台建设关键技术应用,在国家科技文献信息保障体系中发挥引领示范作用。

1.4 开放获取正成为科技文献的重要形态

开放获取(Open Access, OA)正成为未来科技文献工作的一个趋势。公共资金资助科研成果的开放获取已经成为世界各国的共同行动,多数科技强国和科技活跃国家已经出台强制性的OA政策,鼓励公共资金资助科研成果的开放出版,要求公共资金资助科研成果存储到开放知识库实行开放共享。同时,PLOS、BMC等OA出版社已经成为科技论文发布的可信赖重要渠道,多数传统出版社也已大规模出版开放期刊,积极推动订购期刊向开放期刊转变。资助机构、出版社和科研教育机构正积极尝试将图书馆订购经费转换为论文开放出版费用,新的科技期刊运营形态正在酝酿和发展。

NSTL积极组织开展我国OA政策措施研究和实施路径设计,主动进行OA实践探索。2014年,经科技部批准,NSTL代表我国参加高能物理开放出版资助联盟(SCOAP3),履行中国在其中的相关权利、责任和义务,支持全球高能物理学术成果的开放获取,对促进我

国科研成果的开放交流,提高我国在国际科技界的影响力与地位发挥了重要作用。2017年10月,NSTL签署了“OA2020”计划倡议的《关于大规模实现学术期刊开放获取的意向书》,表明NSTL积极支持和推动学术期刊从订阅向OA转变的国际化合作,体现了国家平台的积极引领作用。在2018年12月召开的第14届柏林开放获取会议上,NSTL与国家自然科学基金委、中国科学院文献情报中心(代表中国科学院)共同发表立场声明,支持OA2020倡议和开放获取S计划,支持公共财政资助项目的研究论文立即开放获取。目前,NSTL采集相当数量的OA电子资源供全国科研工作者使用,包含OA期刊9 945种、OA会议录5 302种、OA科技报告9 680种、OA学位论文91 091份。

2 我国科技文献保障工作面临困难和挑战

随着中国特色社会主义进入新时代,创新成为引领发展的第一动力,科技自立自强成为国家发展的战略支撑,对科技文献的服务和利用提出更高的要求。当前,科技文献资源的生产、利用与传播已经全面步入数字时代,数字出版、语义出版、社交网络、开放获取正在成为信息资源生产的主流,需要对多源异构数据进行规范整合、关联计算、关系推理和知识发现^[2]。因此,国家科技文献保障体系建设正面临一系列的机遇与挑战。

2.1 新冠疫情影响深远

在全球疫情仍在蔓延的态势下,很多国家经济发展受限,GDP下降,财政收入放缓,科教、文化等预算和公共支出减少,大部分图书馆和信息所的经费在2021年或是负增长或者零增长,过紧日子成为常态。疫情防控常态化对科技文献特别是文献服务工作带来很大影响,国内外交流受限,外文文献采购特别是到货率受到影响;疫情防控对图书馆服务方式和用户文献获取方式形成新挑战、新要求,图书馆和信息所开放受限,不能完全开放,预约制带来困难,读者和用户不能完全自由查询、阅读;文献服务宣传推广活动受到很大限制,无法举办大规模的现场宣传培训会议,网络宣传培训的效果与面对面培训存在差距;受会议、差旅限制,用户调研活动较为困难,无法深入基层和科研一线听取用户意见,获取用户需求。疫情防控常态化需要我们不断

地创新科技文献服务方式,不断加强网络服务和远程服务能力,开拓网络宣传培训的新模式新方法,培育用户的信息素养和能力,将新冠疫情对文献信息服务的不利影响降到最低。

2.2 文献保障不平衡、不充分的问题突出

当前,我国不同地区、不同系统间的科技文献需求和供给水平差异较大。由于国外科技文献资源价格基本由国外供应商垄断,我国一些地区因经费不足、渠道不畅或数字资源利用率不高等原因,较少购买国外科技文献资源,文献整体保障水平极不平衡。从地区来看,我国经济发达地区和经济欠发展地区的文献保障存在较大差距,中小微企业、民营企业以及地方院所很难获得文献保障和服务的支撑。高校图书馆文献资源购置也不平衡,重点高校购置费每年馆均值是普通高校的3倍以上,是高职高专学校的11倍。国家平台优势发挥不够,向公共服务平台开放馆藏资源数据,带动各地区、各行业的科技文献信息服务有很大潜力。科技文献供给不充分,原始创新文献集成不够、加工精度不高,知识库建设薄弱,无法释放文献大数据的真正价值,缺乏基于文献数据的分析型数据库和工具等,产品层次和服务功能不强,亟待强化信息深度加工处理、发展有自主产权的信息工具和平台^[3]。另外,缺乏对主体用户的研究,真正了解用户需求不够,难以实现智能化推送。因此,虽然近年来我国科技文献的保障能力稳步提高,但不同地区、不同领域、不同创新主体的科技信息需求与供给不平衡不充分的矛盾依然突出,亟需应对和解决。

2.3 科技文献保障安全风险形势严峻

严峻复杂的外部环境给国外科技文献的引进造成新的困难。随着我国自主创新能力的不断提升,国外对我国科技信息限制日趋明显,科技文献保障的国际形势和安全风险日益严峻。2018年以来,美国定向能协会会议录、美国政府四大科技报告等重要文献不再向我国销售,部分中国科研机构被美方列入出口管制“实体清单”,禁止其使用美国信息产品,部分美国数据库对名单中的中国机构、企业和学校停止服务,我国科技信息安全面临新的挑战。

2.4 电子资源和数据库受制于人

随着科技文献从印本向数字资源的转变,网络 and 数字文献(资源)采购的仅仅是“使用权”而非“拥有权”,数字信息资源的国家拥有和保存成为重大问题。资源存放在国外,一旦发生自然灾害、局部战争、国际纷争等不可抗力事件,我国对外科技文献访问和使用将遭受阻碍。专业数据库严重受制于人,特别是生物科学、医学、化学等领域问题更加突出。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,有可能暴露我国最新科研动态和研究方向。我国缺乏高质量的科技期刊、会议录等自主知识产权产品,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需要加强数字资源的建设和保障,有计划、有步骤地实现对数字资源、网络资源、OA资源等的完整采集和本地保存。

3 国家科技图书文献中心“十四五”规划

3.1 编制背景

2000年6月,经国务院领导批准,科技部联合财政部、原国家经贸委、农业部、卫生部、中国科学院等正式组建了国家科技图书文献中心,由中国科学院文献情报中心、中国科学技术信息研究所、机械工业信息研究院、冶金工业信息标准研究院、中国化工信息中心、中国农业科学院农业信息研究所、中国医学科学院医学信息研究所、中国标准化研究院标准馆和中国计量科学院文献馆9个文献信息机构组成。NSTL以构建数字时代的国家科技文献资源战略保障服务体系为宗旨,坚持联合、协作、开放、共享的原则,按照“统一采购、规范加工、联合上网、资源共享”的运行机制,采集、收藏和开发理、工、农、医各学科领域的科技文献资源,面向全国提供公益性、普惠性的科技文献信息服务。经过20年的发展,NSTL建立了印本文献战略储备和数字资源国家保障并举的科技文献资源体系,印本文献的采集数量和数字资源的国家保障规模居国内首位。NSTL构建了辐射全国的科技文献信息服务体系,在29个省、自治区、直辖市建立了服务站,构建了面向高校和科研院所的用户管理平台,开通了面向集团用户的接口服务,提升了地方和行业科技文献信息保障水平,成为国家科技文献信息资源保障基地、国家科技文献信息服务集

成枢纽。

NSTL非常重视发展规划和发展战略的研究,先后制定了“十一五”“十二五”“十三五”“十四五”发展规划,20年间已经编制了4个五年规划。在我国科技自立自强发展、强化国家战略科技力量的新发展阶段,科技文献信息保障基础设施和保障服务能力正成为改善科研创新环境、提升创新驱动发展能力的重要手段。NSTL深刻认识国家对科技信息服务的战略需求和国内外科技信息服务的范式演变,努力抓住科技服务发展机遇期和信息环境重大演变期相互叠加的机会,于2019年启动了“十四五”规划的调研和编制工作,2020年完成初稿后,多次广泛征求业界专家、成员单位意见,反复修改,不断完善,已于2021年1月正式印发。

3.2 发展目标

“十四五”规划确定的NSTL总体发展目标是科技文献信息资源保障能力显著增强,基本实现不同地区、不同系统、不同创新群体间科技文献信息保障的均衡化;科技文献信息服务水平和智能化知识服务能力显著提升,建成以科技文献大数据为核心的、具备知识服务基本功能的国家科技文献信息战略保障体系。具体包括加强国家科技文献本土化保障基地建设,稳固印本科技文献的重要保障,大幅提升数字文献资源协同保障能力,提升科技文献长期保存规模,提高安全保障能力^[4]。建设国家科技文献元数据库,强化知识化处理能力,提高科技文献信息元数据的知识化加工能力,成为支撑国家创新发展的数据挖掘和知识组织推动中心。强化系统和网络基础设施,完善基于大数据的资源发现系统的服务功能,提高发现和知识服务能力。完善国家科技文献信息公益服务网络,提升科技文献的均衡保障能力,增强情报服务能力,构建多层次的情报服务产品。加强关键技术的应用研究,提高可持续发展能力,全面实现NSTL数字业务流程再造,提高引领行业发展的关键技术应用研究和产品研发能力。

3.3 重点任务

“十四五”规划确定了NSTL未来发展的五大任务。

一是建设安全的国家科技文献本土化保障基地。坚持印本和数字文献相互协同的资源保障建设,巩固现有印本文献战略储备,不断调整和优化印本文献学

科结构,加强关键研究领域、前沿技术领域等创新驱动发展需要的资源建设。大力推动数字文献资源的国家统筹机制建设,继续以“国家许可”方式订购全文现刊数据库、回溯期刊全文数据库。加强开放资源的建设和集成,继续采集和存储国际上高质量的OA期刊、会议录、学位论文、图书、课件、机构知识库等资源。增加智库报告、年鉴、进展、综述、述评等三次文献的采集,规划新媒体、科学数据、可视化图谱等新兴资源建设,适时推进富媒体学术资源、事实型数据资源、术语型组织资源、软件型工具资源等资源建设,形成媒体融合时代的立体化资源结构。加快推进国家数字资源长期保存系统建设,实现我国数字资源、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,建立异地备份机制和可靠的公共服务触发管理机制,增强抵御信息风险的能力,保障国家信息安全。

二是全面实施科技文献元数据战略。全力构建国家元数据库,大力拓展元数据多来源获取渠道,加大力度汇聚集成网络数据库、开放资源、回溯数据库的元数据。加强科技文献数据的知识化和语义化组织,深化科技文献的多粒度知识内容描述、概念实体识别和语义关系揭示。继续深入开展名称规范、自动分类标引等知识化加工工作,探索对文献目录、章节知识点、内容提要、实体、表格、图片、公式、概念等的细粒度加工和关联,构建高附加值资源库。推进文献资源的深度聚合、关联计算、科研实体语义识别,按照统一的语义框架组织语义关系库,逐步形成一体化、可计算的开放知识库。

三是强化系统和网络基础设施。完善科技文献发现和服务系统建设,研发包含数十亿级科技文献信息的大数据发现云平台,形成统一集中的科技文献资源聚合服务系统,提供从信息搜索到全文获取的一站式公益性服务。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文献资源和不同服务方式之间的开放动态链接,为用户提供情景敏感的科技文献获取服务。持续优化NSTL计算机网络基础设施的计算、存储、带宽和服务能力,全面支撑核心业务运营、信息服务、关键计算业务、数据存储和备份等重要任务。

四是完善国家科技文献信息普惠服务和情报服务网络。拓展优化科技文献信息普惠服务,优化NSTL服务网络的布局,吸纳更多有条件的行业或地方机构加入NSTL服务网络。基于用户信誉扩展多层次服务和快

速响应策略,深化移动服务,将资源与服务嵌入用户业务流程和业务系统中,提高普惠服务的辐射范围和延伸能力。提高情报服务能力,面向国家政府部门开展战略情报服务,支持国家重大战略决策。继续面向国家战略、重大项目、重点领域开展全流程信息跟踪与专项服务,加强对战略性科学计划和科学工程、国家重点实验室、综合性国家科学中心和区域性创新高地的知识化学科服务,拓展面向企业科技创新的信息服务,建设行业专题平台。

五是探索科技文献智能知识服务关键技术。研究从科技文献、科学数据等数据源中进行数据集成、知识表示、语义关联和用户认知模型构建的方法,以及大规模语义知识库自动构建、知识计算、知识服务、可视化交互等关键技术。重点探索文献知识内容描述、概念实体识别及语义关系抽取、知识计算技术、自动分类综述、跨语种跨媒体关联融合等方法和技术,面向学科评估、情报分析、技术预见等,利用可泛化的知识学习与计算工具引擎,为不同创新需求与功能场景的数据分析评价服务提供泛在的技术支撑。推进知识组织体系的应用,支持国家科技文献平台升级为开放知识应用平台。

4 结语

面向未来,NSTL将继续秉承共建共享的理念,在体制、机制和服务创新上下功夫,以“十四五”规划为统领,紧密围绕规划确定的五大任务,不断提升新时代科技文献信息保障能力和知识化服务水平,引领和促进我国科技文献信息事业发展,为我国科技文献信息保障体系建设与发展做出更大贡献,为我国实施创新驱动发展战略提供更强大的信息支撑。

参考文献

- [1] 新形势下加强基础研究若干重点举措[EB/OL].[2021-04-12].
http://www.cas.cn/zcjd/202005/t20200512_4745744.shtml.
- [2] 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.
- [3] 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.
- [4] 彭以祺.传承发展续写辉煌——隆重纪念国家科技图书文献中心成立二十周年[J].数字图书馆论坛,2020(7):1-2.

作者简介

彭以祺，男，国家科技图书文献中心主任。

Situation of the National Scientific and Technological Literature Guarantee Work and the 14th Five-Year Plan of NSTL

PENG YiQi

(National Science and Technology Library, Beijing 10038, China)

Abstract: Scientific and technological literature is an important scientific research condition and strategic resource. Based on a brief overview of the current situation of scientific and technological document guarantee work in China, this paper analyzes the difficulties and challenges and proposes the 14th Five-Year Plan of National Science and Technology Library. The overall development goal of NSTL is to significantly enhance the ability to guarantee scientific and technological literature information resources, and to improve the service level of scientific and technological literature information and intelligent knowledge service capabilities. The five tasks of NSTL in the future development are determined in the plan, which are to build a safe national science and technology literature localization guarantee base, fully implement the science and technology literature metadata strategy, strengthen the system and network infrastructure construction, improve the national science and technology literature information inclusive service and information service network, and explore the key technologies of science and technology literature intelligent knowledge service.

Keywords: National Science and Technology Literature Guarantee; NSTL; Information Service; Knowledge Service

(收稿日期: 2021-04-15)

■ 书 讯 ■

《汉语主题词表》

《汉语主题词表》自1980年问世以后，经1991年进行自然科学版修订，在我国图书情报界发挥了应有作用，曾经获得国家科学技术进步二等奖。为适应网络环境下知识组织与数据处理的需要，由中国科学技术信息研究所主持，并联合全国图书情报界相关机构，自2009年开始进行重新编制工作，拟分为工程技术卷、自然科学卷、生命科学卷、社会科学卷四大部分逐步完成。目前工程技术卷和自然科学卷已出版。

《汉语主题词表（工程技术卷）》共收录优选词19.6万条，非优选词16.4万条，等同率0.84，在体系结构、词汇术语、词间关系等方面进行了改进创新。《汉语主题词表（自然科学卷）》共收录专业术语12.4万条，包含数学、物理学、化学、天文学、测绘学、地球物理学、大气科学、地质学、海洋学、自然地理学等学科领域，收词系统、完整，语义关系丰富、严谨，每条词汇都有相应的学科分类号表现其专业属性，并与同义英文术语对应。同时，建立《汉语主题词表》网络服务系统，提供术语查询、文本主题分析、知识树辅助构建等服务。《汉语主题词表》可用于汉语文本分词、主题标引、语义关联、学科分类、知识导航和数据挖掘，是文本信息处理及检索系统开发人员不可或缺的工具。

《汉语主题词表（工程技术卷）》已于2014年由科学技术文献出版社出版，分为13个分册，总定价3 880元。

《汉语主题词表（自然科学卷）》已于2018年5月由科学技术文献出版社出版，分为5个分册，总定价1 247元。两卷均可分册购买。