新时期需要强化我国科技信息资源建设

刘细文[1,2]
(1. 中国科学院文献情报中心，北京 100190；2. 中国科学院大学，北京 100039)

摘要：国际科技竞争日趋激烈，已经影响了学术信息交流和信息资源获取。在数字化、智能化时代，需要进一步认识科技信息资源的重要性，理解科技信息资源的内涵与价值，创建新型学术信息生态，强化学术信息资源的安全治理。

关键词：科技信息资源；学术信息生态；数据安全
中图分类号：G35；G25　DOI：10.3772/j.issn.1673-2286.2022.06.002
引文格式：刘细文. 新时期需要强化我国科技信息资源建设[J]. 数字图书馆论坛，2022 (6)：10-13.

当今世界正在经历百年未有之大变局，呈现世界多极化、经济全球化、社会信息化、文化多样化的特点。同时，世界面临的不稳定和不确定因素正在增加，全球经济低迷，单边主义、保护主义抬头，网络安全、重大传染性疾病、气候变化等非传统安全威胁持续蔓延，国际秩序和全球治理体系受到冲击。

党的十九大报告向全社会发出了建设社会主义现代化强国的总号召，并详细阐释了我国强国建设的基本内涵，绘制了制造强国、质量强国、科技强国、文化强国、教育强国、人才强国、网络强国、体育强国、交通强国的发展蓝图。归根结底，强国蓝图的实现还需要进一步锻造国家科技创新能力，实现高水平的科技自立自强。

创新是迎接和应对复杂竞争和挑战的根本手段，通过共享创新成果、加强信息交流，可以促发创新灵感，携手应对共同挑战。数字环境下科技创新更加依赖多样丰富的信息资源和灵敏高效的知识发现工具，更加依赖融合在科技信息资源与工具之中的学术交流。

1 认识新时期科技信息资源的重要性

数字化、网络化、信息化的全面发展，云计算、智能技术、数字技术的全面应用，数字经济、数字产业、数字政府、数字社会的快速进步，让数字信息资源成为社会经济发展的核心要素，同时又为云计算、人工智能等创新发展奠定坚实基础。世界各国都将数字经济模式建设置于优先地位，提出和制订了数字政府、数字经济、智慧城市等系列国家战略规划，创建数字社会的基本架构和管理体制，如《2030数字罗盘：欧盟数字十年战略》（*2030 Digital Compass: the European way for the Digital Decade*）、《数字政府：建立一个面向21世纪的平台更好地服务美国人民》（*Digital Government: Building a 21st Century Platform to Better Serve the American People*）、日本超智能社会5.0（Society 5.0）等。在制订和实施国家数字战略中，这些国家和地区普遍将数据监管置于战略政策的核心位置，如美国国立卫生研究院颁布《数据科学战略计划》（*NIH Strategic Plan for Data Science*），美国国防部发布《数据战略》（*DOD Data Strategy*）。此外，欧盟颁布《通用数据保护条例》（*General Data Protection Regulation*）、《数字服务法》（*Digital Service Act*）、《数字市场法》（*Digital Market Act*）以及英国出台《数字宪章》（*Digital Charter*）等，将数据与国家安全、个人隐私等关联，制定严密的数据保护规则和运营制度。

数据驱动的科研模式已经成为现代科技创新的基本范式，对于科研数据、计算模型、学术信息的认知和应用更加深入。2007年，图灵奖得主吉姆·格雷（Jim Gray）提出了科学研究的实验范式、理论范式、仿真范

式之外新的科研范式，即数据密集型科学发现（Data-Intensive Scientific Discovery）范式。2009年，微软公司主持编制了《第四范式：数据密集型科学发现》，完整展示了数字化科研模式变革的内涵[1]。数据密集型科研范式的变革，催生了诸多学科信息学方法诞生，改变了实验数据分析模式，创建了科研新流程，建立新的科研工具，极大提高了科研效率。2016年5月，Nature发表研究文章[2]，利用机器学习从科研“失败”数据中“学习”，并对新材料进行预测，机器预测结果成功率为89%，而化学家的预测判断成功率为78%。2020年，美国桑迪亚国家实验室（Sandia National Laboratories）开发机器学习算法，以比正常速度快近4万倍的速度进行材料模拟计算[3]。2021年7月，DeepMind宣布AplhaFold2成功预测98.5%的人类蛋白质结构，其预测的所有氨基酸残基中，有58.0%达到可信水平，其中有35.7%达到高置信度[4]。而在此之前科学家花费数十年的努力，也只是认识了覆盖人类蛋白质序列中17%的氨基酸残基。所有这些机器学习都是充分利用了已有的科技信息。

因此，在数字经济时代，我们需要进一步提高对科技信息资源重要性的认识。第一，需要认识科技信息资源是数字经济发展的核心要素和关键，也是国家实力的体现。第二，科技信息资源的范围不断扩大和丰富，不再局限于学术交流的科技文献，科研数据、学术论文、工具软件、学术交流平台等也成为科技信息资源的重要组成部分，同时，科技信息资源的范围也不再局限于自然科学和技术领域，社会科学领域的数字化信息资源大量涌现，体现了自然-社会-人文学科交叉融合的特点[5]。第三，要将数据安全、数据隐私、数据伦理、数据市场监管等纳入数字治理内涵，要将数据治理置于国家安全、数字经济发展中的重要地位。

2　深刻理解科技信息资源的内涵与价值

信息资源是数字经济的核心要素，是社会财富，也是国家软实力和竞争力的重要标志。信息资源既可以直接创造商业产品价值，也可以与传统产业融合，改造传统流程，优化资源配置，间接产生对物质资源和能量资源的替代效应，提高附加值。因此，数字产业既是数字经济的重要产业部门，也是数字经济发展的基石。

科学技术是第一生产力，科技信息资源更是生产力的倍增器和放大器，可以提高科学研究的效率和效用，拓展科技知识的应用范围。在科研数字化时代，科技信息资源贯穿科技创新的全过程，创建新型科研模式，构建新的科研工具和方法，建立新型科学认识模型和框架，实现高效的科技创新。2021年，IEEE在其年度趋势预测图景中，展示了2025年在科技创新“长河”中发挥作用的科研设施、学术信息流、科研机构、出版传播、信息保存、信息评价等。在学术信息交流过程中存在多重形态的科技信息资源，信息和数据累积到一定程度后形成资源，可以有效展示信息内部的结构关系，从而具备知识生产要素的属性[6]。我们可以将科技信息资源分为数据、论文、学术交流信息、社会交流信息、科研条件信息等若干层次，展示出科技信息资源的框架结构。在数据层面，有科研观测原始记录数据、规范整理的科研数据、支撑学术思想交流的数据、非正式学术交流的数据化信息等；在论文层面，有同行评议学术论文、半同行评议论文和非同行评议论文，从论文形式看，有期刊论文、会议论文、研究报告、预印本论文、学位论文等；在学术信息交流层面，有学术活动信息、学术观点传播、学术争鸣、学术评价信息等；在社会交流信息层面，有科技新闻信息、社会传播信息、个人自媒体信息、社会评价信息等；在科研条件方面，有科研机构信息、科研人员信息、科研仪器设备信息、科研工具软件等[7]。各类信息共同构成一个完整的科研信息生态，形成科技信息资源体系。

随着数字社会建设的飞速发展，科技信息资源内涵的不断丰富，让科技信息在科技创新中的作用不断得到加强，其价值和效用得到了较大体现。科技信息资源与科技创新活动形成了双向驱动的增益循环，让科技创新活动成为数字化的知识生产过程。科技信息的不断生产和流通，形成了数字化学术信息交流产业；同时，科技信息资源对科技创新活动发挥助推作用，助力生产更加丰富的学术信息。

3　重视创建数字化学术信息生态

数字经济场景下和数字科研发展趋势中，数字化学术生态成为科技创新基础设施的重要组成部分，决定了科技创新能力和效率。建设与数字经济生态相协调的数字化学术信息生态，是充分发挥数字信息资源服务效能的必然要求。

2021年底，我国发布了《“十四五”数字经济发展规划》[8]，将加强数字生态建设作为重要任务之一，通

过建立有效的制度环境，以法治为基础，坚持促进发展和监管规范两手抓，着力构建数字生态规则体系，全面提升数字生态治理能力，推动数字生态健康、有序、可持续发展。数字生态建设的内涵丰富，包含数据要素、基础设施、数字产业、数字服务、数字监管和治理等方面。在数据要素生态方面：需要建立可以不断丰富数据类型和数字资源的规则，进一步完善数据资源开放共享制度，推动全社会各类数据良性互动、融合应用，打破“数据孤岛”，释放数据红利；推动数据交易市场建设，让全社会深度挖掘数据要素价值，释放以数据驱动经济社会发展的强劲动能；建立健全数据安全管理、风险评估、检测认证等机制，完善适用于大数据环境下的数据分类分级保护制度，加强对海量数据汇聚融合的风险防护，强化数据资源全生命周期的安全管理。在基础设施生态方面：搭建多技术融合的网络信息结构，形成支撑数据生产、数据采集、数据汇聚、数据存储、数据计算、数据分发、数据决策的设施能力体系，让数据设施与数据生态形成互动。在数字产业生态建设方面：创建形式多样、服务能力不断优化的数字要素加工、组织、集成、创意、推送、互动的数字产业生态，形成数字技术产业、数字内容产业、数字服务产业全流程格局，建立数字产业与实体产业的互动型发展格局。在数据服务生态方面：加强互联网内容建设，充分发挥数字化、网络化、智能化传播优势，建设文明网络、开放网络、公平网络、安全网络、便捷网络，丰富全民的数字化生活场景和体验，打造智慧共享、和睦共治的数字生活新模式，充分利用数字技术，强化数字监管流程，创建数字监管规则。

同样，数字化学术信息生态建设需要从科研数据要素、创新工具、创新过程、创新方法、创新平台等多角度着手，全方位深入理解学术信息生态价值和作用，多途径快速推进实施。数字化学术信息生态是由科学数据、学术文献、交流渠道、交流工具、交流平台等协同构成。在开放科学背景下，学术信息生态需要以建设科学数据基础设施、文献与知识服务设施为抓手，促进科学数据、学术文献等通过物理或虚拟网络汇聚、关联、存储、应用，让学术信息资源发挥规模化新型科研生产资料的作用。例如，发展以AlphaFold2为代表的“智能科学家”机器人、通用/专业软件工具，在数字环境下规范、高效地实践科研构想，满足远程控制、实时分析、分布式协同交互、个性化研究需求，并将新的科学数据资源、知识以数字化形态融入科研生产资料中。学术信息生态还需要通过加快建设智能化学术实验室，在提供一站式、全景式的数字化科研环境的同时，构筑服务研究人员知识传播的高端交互场所，为打破传统学科领域边界的学术思想、技术理念的碰撞提供平台。学术信息生态建设，需要充分借助数字化形态资源条件升级自然-社会-人文科学交互实践，从搭载电子馆藏、数字化期刊资源，到集成虚拟学术社区，在数字学术空间下催生更加多源、开放、深度交互的学术交流新模式。以科学数据管理和知识交互协同为核心的数字化学术生态，将为应对科技前沿与社会经济发展挑战的跨学科、系统性、综合性问题提供新的路径和机遇。

4 健全科技信息资源安全治理

学术信息资源是提升国家竞争实力的关键要素和战略资源。科技信息资源、科学数据的安全受到各国普遍关注。科技信息资源的安全需要遵从普遍意义上的数据安全规则。数据安全有两方面的含义：一是数据本身的安全，主要是指采用现代密码算法对数据进行主动保护，如数据保密、数据完整性、双向强身份认证等；二是数据防护的安全，主要是采用现代信息存储手段对数据进行主动防护，如通过磁盘阵列、数据备份、异地容灾等手段保证数据的安全。

近年来，全球数据安全的相关法律法规相继出台，各国加快步伐织密数字经济“保护网”。美国采取分行业分散式立法模式，颁布了一系列的数据保护联邦立法，在电信、金融、医疗健康、教育以及儿童在线隐私等领域设有专门的数据立法保护。欧盟践行统一立法模式，由《通用数据保护条例》、《非个人数据自由流动条例》（*Regulation on the Free Flow of Non-personal Data*）、《电子隐私条例》（*Regulation on Privacy and Electronic Communication*）3部法律统率数据领域，2022年密集出台《数据服务法案》（*Digital Service Act*）、《数字市场法案》（*Digital Market Act*）、《数据治理法案》（*Data Governance Act*）、《数据法案》（*Data Act*）等一系列法律政策文件，从数据流转、开发利用、促进良性竞争、基本权利保护等侧面勾勒数字经济安全保障欧洲方案。我国颁布《数据安全法》和《个人信息保护法》，从保护国家安全和公共安全、保护公民隐私角度开展数据安全治理，与《网络安全法》《保密法》等共同构建国家数据安全法律框架。为顺应数字经济发展，满足合规成为创新主体适应数字经济发展的必

要条件和基本需求，要在组织规范和技术层面完美整合，实现体系化数据安全，让安全从成本变成竞争力。

科技信息资源安全除涉及数据本身的自然风险、技术风险外，还包括其承载的利益相关者的权益与利益安全，且受到社会环境、技术水平、数据类型和规模等影响，呈现敏感性、时效性、动态性和传动性。如头部学术出版商通过收购逐步扩大数据业务版图，垄断之势带来信息资源保障风险。15家具有全球影响力的学术出版机构向俄罗斯“断供”科学和知识产权信息服务，美国政府严格审查和限制与中国的科技交流，禁止列入实体清单的机构（如哈尔滨工业大学、哈尔滨工程大学）使用MATLAB软件，这些事件时刻提醒我们科技信息数据也面临政治风险。因此，必须从法律政策、管理、技术与平台等维度采取多种数据安全治理措施，构建科技信息安全治理体系，提高我国科技信息资源治理水平与国家治理能力。

国家相关部门应进一步高度重视科技信息资源的安全问题，将学术信息资源安全与开放共享同等考虑，在国家总体安全战略层面审视学术信息资源的共享与安全。首先，要加快出台科技信息资源、平台、工具安全专项法律政策，切实保障国家科技信息主权，维护自身科技信息与数据的管控权、对外平等独立的处理权以及相关核心软件工具、硬件设施的自主权。其次，要依托科学数据基础设施开展常态化安全治理，开展科技信息资源的普查，摸清虚拟空间我国科技信息资源、平台、工具状况。最后，要开展科技信息安全理论与战略研究，明确数据主权安全红线，绘制科技信息数据“疆域”，定立主权“边界”，探索本地备份出境数据机制，监测数据流向并定期进行风险预警分析，有效防止科技信息数据流失。

参考文献

[1] HEY T, TANSLEY S, TOLLE K. 第四范式：数据密集型科学发现 [M]. 潘教峰，张晓林，译. 北京：科学出版社，2012.

[2] RACCUGLIA P, ELBERT K C, ADLER P D F, et al. Machine-learning-assisted materials discovery using failed experiments [J]. Nature, 2016, 533: 73-76.

[3] DE OCA ZAPIAIN D M, STEWART J A, DINGREVILLE R. Accelerating phase-field-based microstructure evolution predictions via surrogate models trained by machine learning methods [J/OL]. npj Comput Mater, 2021 [2022-06-01]. https://doi.org/10.1038/s41524-020-00471-8.

[4] TUNYASUVUJAKOOL K, ADLER J, HASSABIS D, et al. Highly accurate protein structure prediction for the human proteome [J]. Nature, 2021, 596: 590-596.

[5] 刘细文. 中国科学院文献情报中心“十四五”发展思考——基于数据、信息、知识与情报的规划框架设计 [J]. 数字图书馆论坛，2021（5）：12-16.

[6] STM. STM Trend 2025 [EB/OL]. [2022-07-01]. www.stm-assoc.org/standards-technology/stm-trends-2025/.

[7] 刘细文. 情报学范式变革与数据驱动型情报工作发展趋势 [J]. 图书情报工作，2021，65（1）：4-11.

[8] 国务院印发《“十四五”数字经济发展规划》[BE/OL]. [2022-06-01]. http://www.news.cn/politics/2022-01/12/c_1128256301.htm.

作者简介

刘细文，男，1965年生，博士，研究员，中国科学院文献情报中心主任，研究方向：情报学。

Strengthening the Construction of S&T Information Resources in the New Ages

LIU XiWen[1,2]

(1. National Science Library of Chinese Academy of Sciences, Beijing 100190, P. R. China;
2. University of Chinese Academy of Sciences, Beijing 100039, P. R. China)

Abstract: The increasingly competition in science and technology in the world has affected seriously the exchange of academic information and the acquisition of information resources. In the digital and AI era, we need to further understand the importance of scientific and technological information resources, understand the connotation and value of scientific and technological information resources, create a new academic information ecosystem, and strengthen the security management of academic information resources.

Keywords: Scientific and Technological Information Resources; Academic Information Ecosystem; Data Security

（收稿日期：2022-06-10）