

引用格式:杨雅萍,姜侯,孙九林.科学数据共享实践:以国家地球系统科学数据中心为例[J].地球信息科学学报,2020,22(6):1358-1369. [Yang Y P, Jiang H, Sun J L. Practice of scientific data sharing: A case study of the national earth system science data center[J]. Journal of Geo-information Science, 2020,22(6):1358-1369.] DOI:10.12082/dqxxkx.2020.200110

科学数据共享实践:以国家地球系统科学数据中心为例

杨雅萍^{1,2,3}, 姜侯^{1,4}, 孙九林^{1,2,3*}

1. 中国科学院地理科学与资源研究所 国家地球系统科学数据中心,北京 100101; 2. 中国科学院地理科学与资源研究所 资源与环境信息系统国家重点实验室,北京 100101; 3. 江苏省地理信息资源开发与利用协同创新中心,南京 210023; 4. 中国科学院大学,北京 100049

Practice of Scientific Data Sharing: A Case Study of the National Earth System Science Data Center

YANG Yaping^{1,2,3}, JIANG Hou^{1,4}, SUN Jiulin^{1,2,3*}

1. National Earth System Science Data Center, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; 2. State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; 3. Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China; 4. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: Earth System Science (ESS) is a comprehensive interdisciplinary discipline, which originates from the study of global climate change and benefits from the progress of remote sensing technology. Now, ESS has entered the era of big data and artificial intelligence technology has played a key role in solving the frontier problems of ESS. Scientific data sharing is essential for the prosperity of science development and utilization of data value. After long-term exploration and practice, sound data management policies and mechanisms, continuous data sharing service system, and diversified scientific data integration modes have been established around the world. Innovative development of data sharing is going on thanks to the progress of sharing theory and ideas, such as the popular "findable, accessible, interoperable, and reusable" FAIR principle and data publishing. China has promulgated policies and regulations at the national level, focusing on promoting the development of national scientific data center, collection and management of scientific data resource from national science projects, as well as data publishing and protection. Combing the experience abroad and the actual situation in China, researchers have built the distinctive classification scheme for ESS data resources and major breakthroughs continuously appear in metadata management, distributed interoperability, big data

收稿日期:2020-03-10;修回日期:2020-04-21.

基金项目:中国科学院战略性先导科技专项(A类)(XDA19020304);国家地球系统科学数据中心(2005DKA32300);中蒙俄国际经济走廊多学科联合考察项目(2017FY101300);中国工程科技知识中心项目(CKCEST-2019-1-4);云南省数据共享发展战略咨询研究(2019YNZH6)。[**Foundation items:** Strategic Priority Research Program of the Chinese Academy of Sciences (A), No.XDA19020304; National Earth System Science Data Sharing Infrastructure, No.2005DKA32300; Multidisciplinary Joint Expedition for China-Mongolia-Russia Economic Corridor, No.2017FY101300; Branch Center Project of Geography, Resources and Ecology of Knowledge Center for Chinese Engineering Sciences and Technology, No.CKCEST-2019-1-4; Data Sharing Design of Yunnan Province, No.2019YNZH6.]

作者简介:杨雅萍(1964—),女,北京人,高级工程师,主要从事地球系统科学数据共享研究。E-mail: yangyp@igsnr.ac.cn

*通讯作者:孙九林(1937—),男,江苏盐城人,研究员,主要研究地球系统科学大数据与人工智能。E-mail: sunjl@igsnr.ac.cn

analysis, scientific data sharing services, and other professional technologies, covering the whole life cycle from data collection, integration, analysis to open and sharing. Taking the National Earth System Science Data Center as an example, we summarize the progress of data sharing services and key technologies, and introduce the practice and achievements in China. At present, the national data sharing work in the field of ESS has contributed towards the formation of mature and stable operation mechanism, established a formal standard framework for multi-source distributed scientific data integration, developed multi-scale geoscience database covering diverse disciplines and themes, and built up distributed service networks and systems suitable for massive heterogeneous data sharing. This work not only promotes the development of Geoscience, but also fastens the dissemination and promotion of data sharing theory. However, issues such as isolated data islands, low-level generalization of service systems, and weak accords with international standards still hinder the advance of data sharing. In the future, with personalized needs for data sharing activated, customized "data + knowledge" services are expected to become the prevailing modes, which will bring new opportunities and challenges to data sharing.

Key words: Earth System Science; data sharing; data resource classification; big data; key technologies; data publishing; scientific project management; National Earth System Science Data Center

***Corresponding author:** SUN Jiulin, E-mail: sunjl@igsrr.ac.cn

摘要:地球系统科学是一门综合性交叉学科,起源于全球变化研究,受益于遥感技术进步,现已进入大数据时代,人工智能技术成为解决地球系统科学前沿问题的新手段。科学数据共享是推动科学发展、提升数据价值的关键,经过长期的探索和实践,全球已形成完善的数据资源管理政策与机制、持续的数据共享服务体系、多元化的科学数据整合模式,并且新的理念不断推动数据共享创新发展,其中“可发现、可访问、可互操作和可重用”原则和数据出版备受推崇。我国颁布了国家层面的政策法规,着重推动国家级科学数据中心建设、科技项目数据资源汇交及数据出版。借鉴国外数据共享经验,结合国内实际发展情况,我国探索出具有中国特色的地球系统科学数据资源分类体系,在元数据、分布式互操作、大数据分析和共享服务等关键技术方面都取得了重大突破。本文以国家地球系统科学数据中心为例,系统梳理了国内地球系统科学领域的科学数据共享实践进展,详细介绍了国家在地球系统科学领域的科学数据共享实践及其成效。国家地球系统科学数据共享工作已探索形成成熟稳定的运行服务机制、研制出适用于多源分散地球系统科学数据的标准体系、建成多学科多时空尺度的地球系统科学数据库群、自主研发海量异构数据共享网络技术体系与服务系统,促进了地球系统科学的科学发展,同时也推动了数据共享理念的传播与推广。然而,数据孤岛现象、共享服务系统标准化程度低、系统技术标准未对接国际规范等问题仍然较为突出,阻碍了地球系统科学数据开放与共享。未来,随着共享个性化需求被激活,定制化的“数据+知识”服务有望成为主流趋势,将给数据共享带来新的机遇和挑战。

关键词:地球系统科学;数据共享;数据资源分类;大数据;关键技术;数据出版;科技数据汇交;国家地球系统科学数据中心

1 引言

地球系统科学可追溯至20世纪的全球变化研究,1983年美国国家航空航天局(NASA)首次明确提出地球系统科学的概念,1988年NASA出版“Earth System Science: A Closer View”^[1]科学研究报告,标志着“地球系统科学”的起步。出于理解的不同,地球系统科学的基本内涵有所差异,德国科学家 Schellnhuber^[2]认为依靠遥感技术从空间观测地球的结果就是地球系统科学;汪品先^[3]认为地球系统科学是研究地球上各种现象和过程的各独立学科在系统论思想指导下相互结合、融合发展形成的

整合性学科;周秀骥^[4]认为地球系统科学是研究地球系统整体的结构、特征、功能和行为的科学,采用复杂系统科学理论和方法以及现代高新技术手段研究地球系统演变规律、调控机制及变化趋势,阐述自然和人类活动与地球系统变化的相互作用。

自兴起以来,地球系统科学不断在基础理论、研究方法、观测系统、科学研究等方面获得发展和突破^[5-6],其各个分支学科深度渗透融合,促使人们面向整个地球系统开展新的研究^[7];卫星遥感、空间定位、网络通讯、云计算等技术成为地球系统时空立体观测、海量数据分析的强有力工具^[8]。目前地球系统科学正处在全新的发展时期,在原始数据获

取上,逐步形成空天地一体化的现代过程观测体系;在系统模拟和预测方面,超级计算机使得高度复杂非线性地球系统的模拟成为可能;与云计算、人工智能等新技术的融合推进着地球系统科学大数据进程^[9]。遥感卫星源源不断地为人类提供超高维度和超高频次的地球观测数据,催生地球系统科学大数据时代^[10]。地球系统科学数据量正呈现爆炸式的增长,初步估计当前地球系统数据量已经超过100 PB,并且以每年5 PB以上的数据量在增长,已具备大数据的5V(Volume、Variety、Velocity、Value和Veracity)典型特征^[11]。

在大数据时代,科技创新越来越依赖于科学数据的分析挖掘和综合利用,实现科学数据共享有利于促进创新、发挥数据价值、实现科学可持续发展^[12]。地球系统科学数据是国家重要战略资源,加强数据资源规范化管理、有效推动数据共享对于促进人类对地球本身的科学认知、增强人类适应全球环境变化的能力、服务于区域和全球可持续发展具有重要意义。本文梳理了国内外科学数据共享发展现状,重点探讨了地球系统科学领域内的数据共享服务及其关键技术研究进展;并以国家地球系统科学数据中心为例,详细介绍了国家在地球系统科学领域的数据共享实践及其成效;最后结合当前问题和应用需求,展望未来的地球系统科学数据共享发展趋势。

2 科学数据共享进展

数据共享是一项系统性的复杂工程,涉及从科学数据获取、集聚、保存、加工到发布共享和应用的整个生命周期^[13-14],研究内容包括科学数据标准规范、分类体系、管理模式、共享机制等。发达国家数据共享工作起步较早,如美国最早在1960年就成立国家大气研究中心负责地球科学数据的建模、收藏和保存工作^[15];1995年英国经济和社会研究委员会出台数据管理计划,要求对科学数据进行质量管理以实现长期保存和对外共享^[16]。经过多年的发展,国内外科学数据共享均已取得突破性进展。

2.1 全球发展现状

2.1.1 形成完善数据资源管理政策与机制

美国及欧盟等发达国家已形成从宏观到微观的全方位数据共享政策法规,宏观政策解决全局性、长远性和根本性的问题,微观政策解决局部的、

当前的、具体的问题^[17];例如美国国会通过《开放政府数据法案》,推动政府数据开放共享^[18];欧盟发布《通用数据保护法案》,为公民隐私数据保护提供法律保障^[15];法国政府颁布《数字共和国法案》,保障互联网用户对数据的无偿获取和使用,并且强调用户个人隐私保护^[19]。世界数据系统(WDS)、国际科技数据委员会(CODATA)、地球观测组织(GEO)等国际组织对地球系统科学数据的管理和共享极为重视,制定了特定的数据政策与机制保障领域内数据开放与共享;“未来地球计划(Future Earth)”、中国“数字一带一路”等科学工程都制定了科技资源管理规范 and 开放共享计划,并研发地球大数据服务系统保障数据共享服务。

2.1.2 数据管理机构持续开展数据共享服务

国际科学联合会(ICSU)是目前科学界最有权威的非政府国际组织,指导成立了众多科学组织和机构,其中,WDS以确保广泛的数据获取、长期的数据管理和便捷的数据服务为发展宗旨,CODATA在提高数据质量、加强数据管理和推动数据传播方面取得了瞩目的成绩^[20]。美国的国家航空航天局(NASA)和国家海洋和大气管理局(NOAA)、欧盟开放数据平台(ODP)、欧洲数据中心联盟(EUDCA)等国际地学相关科学数据机构,以及世界遥感大气数据中心(WDC-RSAT)、世界土壤数据中心(ISRIC-WDC Soils)等典型地学数据中心,在应对全球变化、陆地表层系统与入地关系及区域可持续发展等研究中持续提供科学数据支撑,取得了显著的效益和国际影响。中国也启动实施了科学数据共享工程,在地球系统科学等基础与前沿领域建成国家级数据共享服务平台,推动国内科学数据资源交换与共享,并逐步打通国际合作渠道。

2.1.3 科学数据整合模式趋于多元化

发达国家依靠权威科学数据组织机构/数据中心,持续整合和汇聚全球科学数据资源,并形成标准化的科学数据收集、管理和存储解决方案。我国既有部门内部“自下而上,纵向整合”模式,也有“跨部门联合,横向整合”模式;既有针对中央财政支持的国家科技计划项目资源的“数据汇交”整合模式,也有针对区域优势资源进行“特色加盟,优势互补”的整合模式;例如,国家地球系统科学数据中心通过“联合共建、资源汇聚、资源交换、项目/论文数据汇交”4种机制,多方位保障科学数据资源的持续收集与整合。

2.1.4 科学数据共享理念不断创新

随着知识产权保护受到重视提高,数据出版成为当前科学数据共享的新举措^[21],各大出版集团和期刊积极研究制定相应的数据出版政策,通过“数据提交、同行评审、数据发布、数据永久存储、数据引用和数据影响评价”等环节发布科学数据,分配统一的资源标识符,推动数据生产者权益保护和科学数据资源价值共享;如CODATA旗下数据科学杂志Data Science Journal、Nature出版集团数据期刊Scientific Data、Earth System Science Data等要求所有提交的数据论文都要经过严格的同行评议,以保证数据的真实性和准确性^[22]。新的科学数据管理理念也不断融入到地球系统科学数据共享,如e-Science数字管理概念、科学数据全生命周期管理体系^[23]、现代科学数据共享“可发现(Findable)、可访问(Accessible)、可互操作(Interoperable)和可重用(Reusable)”FAIR原则^[24]。

2.2 中国科学数据共享

2.2.1 国家推动科学数据管理与共享

紧跟国际科学数据共享发展步伐,我国出台了一系列政策法规保障科学数据管理与共享^[25-26],2004年科技部印发《2004—2010年国家科技基础条件平台建设纲要》着手推动我国科技资源整合共享;次年成立国家科技基础条件平台中心,切实推进科技基础条件平台和科学数据资源的专业化管理;2015年国务院印发《促进大数据发展行动纲要》明确提出发展科学大数据工作,积极推进政府信息系统和公共数据互联开放共享;2018年2月,国家出台《国家科技资源共享服务平台管理办法》,进一步明确各类科技资源共享服务平台的数据共享职责与任务,同年3月国务院办公厅印发《科学数据管理办法》提出科学数据分级管理、安全可控、充分利用的原则和“开放为常态、不开放为例外”的共享理念,标志着我国科学数据管理和共享新的历史阶段^[27]。

2.2.2 建设国家级科学数据中心

为切实推进科学数据开放共享,1999年国家在科技基础性工作和社会公益性研究专项中,启动了科技基础数据库建设。2001年,科技部科学数据共享调研组提出“实施科学数据共享工程,增强国家科技创新能力”的建议,同年底,“气象科学数据共享试点”启动。2002年,国家又启动了地球系统、水文、海洋、地震、国土、农业、林业、人口健康数据共

享,和气象数据共享一并作为国家科学数据共享工程第一批9个试点。此后,试点工程在资源环境、农业、人口与健康、基础与前沿等领域共24个部门全面展开,包括气象、测绘、地震、水文水资源、农业、林业、海洋、国土资源、地质与矿产、对地观测等行业领域国家科学数据共享中心和地球系统、人口健康、基础科学、先进制造与自动化科学、能源和交通等学科领域科学数据共享网。试点工程有效推动了科学数据共享的开展,为科学数据共享的全面建设奠定了扎实的基础^[28]。2019年科技部推进国家科学数据中心建设,对原有国家科技资源共享服务平台进行优化调整,遴选出“国家地球系统科学数据中心”等20个国家科学数据中心,推动建设国家级数据科学创新示范基地^[29]。此外,响应国家需求,中国工程院于2012年启动中国工程知识中心项目,中国科学院于2018年启动地球大数据科学工程建设。

2.2.3 全面开展科学数据汇交与出版实践

2008年7月,国家发布《973计划资源环境领域项目数据汇交管理方案》,并成立“973计划资源环境领域项目数据汇交管理中心”全面负责接收、保存、管理国家“973计划”项目产生科学数据;2011年科技部启动实施国家科技计划项目科技资源汇交与共享工作,进一步整合集成国家科技基础性工作专项科技资源成果,颁布《科技基础性工作专项项目科学数据汇交管理办法(试行)》,推动科技项目数据资源对外开放共享^[30]。在数据出版方面,2014年6月中国科学院地理科学与资源研究所率先创办《全球变化数据学报(中英文版)》,2015年8月中国科学院计算机网络信息中心创办《中国科学数据(中英文版)》,2018年2月全球地球科学领域首个大数据刊物《地球大数据(Big Earth Data)》创刊,推动了我国科学数据出版与数据引用发展^[31]。

3 地球系统科学数据共享关键技术

3.1 科学数据分类

地球系统科学具有明显的学科交叉和区域集成特征,不仅需要地球科学各分支学科基础数据,更需要交叉学科、多尺度、跨区域的综合集成数据^[31],构建科学合理的科学数据分类体系是地球系统科学多类型数据资源管理、整合、集成及高效共享利用的必要前提^[32-34]。国家地球系统科学数据中心基于

国际数据共享机构主题分类思想,并结合在该领域的长期数据共享实践,建立了具有我国特色、同时又能满足数据中心用户需求和数据集成要求的地球系统科学数据分类体系。

廖顺宝等^[32]遵循科学性、系统性与完整性、层次性、可扩展性、实用性以及兼容性的原则,考虑中国地球系统科学数据资源的实际情况形成了地球系统科学的数据资源分类体系,该分类系统以地球为核心,首先向外划分为岩石圈、陆地表层、海洋、大气和外层空间五圈层系统;进一步划分为地质、地震、地球物理、地球化学、大地测量、地理、自然资源、生态、环境、对地观测、海洋、大气、空间和天文14个一级学科分类;最后根据每个学科主要数据内容细分为88个数据类别。王卷乐等^[33-34]结合国家地球系统科学数据中心的用户服务特征,提出了数据目录分类和关键词表分类2种模式相结合的地球系统科学数据分类体系,既满足了数据共享服务过程中的用户需求又兼顾了数据分类的科学性和系统性。随着地球系统科学从深部到海洋和陆表的多圈层探索不断深入、多学科深层次互相渗透和融合、区域协同发展成为全球倡议、全球变化和可持续发展背景下的人地关系研究受到关注,国家地球系统科学数据中心不断优化调整数据资源分类体系,以“圈层+学科”为主线,结合国家发展战略和区域研究需求,形成了涵盖5大圈层、58个三级学科的国家地球系统科学数据中心资源分类体系。

3.2 数据共享服务关键技术

数据共享服务直接面向数据使用者,通过数据查询检索、资源共享等方式将数据(或挖掘产生的知识)提供给目标对象,涉及智能检索、目录服务、数据服务、数据安全等关键技术。

智能检索技术是数据服务的基础,构建地球系统科学数据基于时间、空间和主题内容特征的定量化精准语义关联网络^[35],可实现数据高效检索及面向用户的数据智能推荐^[36]。目录服务技术根据数据资源体系组织元数据集合,面向用户提供数据导航服务,面向计算机提供目录接口服务^[37],开放地理空间信息联盟(OGC)规定了地理空间信息目录服务接口规范,包含发布、发现和管理3类接口。数据服务技术,通过统一数据服务模型,屏蔽多源、异构数据的差异,实现数据统一访问、多层次服务;地球系统科学数据开放技术主要遵守OGC开放数据

服务框架,提供多层次地理信息主题及领域内交叉性主题(如地球影像、遥感反演产品)数据服务。科学数据共享安全近年来受到高度重视,国家《科学数据管理办法》明确强调要保障科学数据安全^[13,27],现有数据服务平台多采用身份认证、访问控制、完整性校验、数据加密、安全审计等技术保障数据存储安全、传输安全以及保护数据版权^[38],最近受追捧的区块链技术也有望为数据共享安全提供新手段^[39]。

3.3 数据收集关键技术

数据收集阶段主要进行原始数据的采集和积累,其核心关键在于元数据处理。元数据技术包括对数据资源的内容、格式、范围、质量、来源、加工流程、使用方式等基本信息的描述和规定,是科学数据共享体系中连接数据生产者、使用者和管理者的纽带^[40]。地球系统科学学科覆盖面广、数据类型复杂多样,一种元数据标准通常很难描述所有类型数据资源;廖顺宝等^[32]认为地球系统科学元数据技术必须支持标准框架下的个性化定制和扩展,覆盖核心元数据标准、元数据语义规则与元数据扩展机制等基础框架;学科或专题元数据标准可基于此面向学科或专题应用需要进行扩展。地球系统科学元数据需重点强调数据时间跨度、空间覆盖范围、获取尺度、空间参考和坐标、时空表达方式以及比例尺和时空分辨率等与时空要素相关的元数据信息,解决好多源、多时相、多尺度、多分辨率和多种类型数据的组织、管理、访问和获取问题^[41],提升数据资源的学科覆盖面和时空完备性^[5,12]。

科学数据汇交是科学数据收集的重要途径,需要从应用层、支撑平台层和数据层构建数据汇交流程体系,对成果登记、数据传输、元数据汇交、数据溯源等基础技术进行规范化^[41-42]。保证数据的真实性和可靠性是数据汇交工作的基本要求,基于数据溯源的数据质量控制与可信管理至关重要;当前行业应用较为广泛的数据溯源框架包括信息描述模型、通用表达模型、领域应用模型、安全管理模型与区块链溯源管理模型等,针对关系数据库、科学工作流、大数据平台、云计算和区块链等典型应用环境的数据溯源计算方法已相对成熟^[43]。

3.4 数据管理关键技术

数据管理是数据共享服务的基础,需重点突破分布式互操作技术。国内科学数据共享平台数据资源建设可分为专业化生产模式、联合共建共享模

式和数据交换模式3类^[44],专业化模式多满足国家层面数据资源保藏需求;联合共建共享模式和数据交换模式都涉及“物理上分布、逻辑上统一”的多节点数据资源的互操作。通过互操作技术,多个数据管理维护单位共同推进数据资源整合工作,为用户提供一站式的数据共享服务、支持个性化的专业数据定制服务,同时能有效解决单点服务带来的网络堵塞、系统负载过大等问题^[37]。地球系统科学数据种类繁多,数据间具有时间相异、基准相异、结构相异、平台相异等多源异构性,需建立多级异构数据交换系统框架,采取中间件技术和动态投影技术等多源异构数据集成子系统,保证不同结构或不同格式数据互操作的协调一致^[45]。

3.5 数据分析关键技术

数据分析旨在从海量数据中获取有效信息,是传统数据服务转型“数据+知识”服务的关键^[46];在大数据时代,数据密集型科学成为科学研究第四范式^[47],对数据分析技术提出了新的技术。大数据预处理更加关注多类型异构数据的辨析、抽取和清洗等操作,数据处理更趋向于分布式非关系型模式,数据融合需解决好异构数据、非结构化数据的整合问题,数据网络挖掘、特异群组挖掘、图挖掘等新数据挖掘技术成为重点。此外,数据密集型科学发现离不开智能软件系统的支撑,当前主流的趋势是基于云平台提供大数据智能分析服务,构建“机器学习即服务”的数据分析服务模式^[48]。结合地球系统科学数据特点,其云平台多关注以下功能特征:海量数据存储与管理,海量数据处理、挖掘和分析运算,复杂异构数据在线处理、分析与挖掘算法,第三方应用系统扩展接口,海量数据时间、空间、对象关联关系的交互可视化;当前较为热门的 Google Earth Engine 提供了很好的应用示范。

4 国家地球系统科学数据共享实践

4.1 国家地球系统科学数据中心

国家地球系统科学数据中心(www.geodata.cn)是国家在地球系统科学领域的数据共享实践主体,其实践历程最早可追溯至20世纪80年代国家国土资源信息系统的研究。2003年,地球系统科学数据共享服务平台作为科学数据共享工程首批9个试点之一启动,开始探索研究分散科学数据共享机制、

标准规范、关键技术等,为地球系统科学与全球变化等研究提供数据共享服务^[49];2005年该平台纳入国家科技基础条件平台,2011年地球系统科学数据共享平台进入运行服务阶段,成为首批通过认证的23家国家科技基础条件平台之一,同时被遴选为6个国家层面的数据平台。2018年国家提出建设一批有重要影响的国家科学数据中心;2019年科技部对原有国家科技资源共享服务平台进行优化调整;原地球系统科学数据共享服务平台调整为“国家地球系统科学数据中心”,并从中遴选产生5个国家科学数据中心(“国家青藏高原科学数据中心”、“国家天文科学数据中心”、“国家空间科学数据中心”、“国家冰川冻土沙漠科学数据中心”和“国家极地科学数据中心”)成为20个国家科学数据中心之一。国家地球系统科学数据中心,持续以地球系统科学领域学科发展与科技创新需求为导向,整合、集成、汇聚领域分散科学数据资源,在保证数据共享安全的基础上,为全球变化、陆地表层、人地关系及区域可持续发展等国家科技创新提供数据支撑。

经过多年发展,国家地球系统科学数据中心逐步实现了我国地球系统科学数据共享从无到有的重要突破,成为分散数据资源整合集成的枢纽、国家科研数据汇交应用示范基地、中国数据共享人才培养高地,成为地球系统科学与全球变化领域科技创新、国家重大战略和区域可持续发展研究的支撑力量,在国内外享有较高知名度和影响力^[46]。

4.2 主要实践成效

4.2.1 形成稳定成熟的运行服务机制

在国家科学技术部指导下,国家地球系统科学数据中心由办公室、专家委员会和用户委员会组成。办公室领导战略研究组、标准规范组、数据建设组、数据汇交组、产权保障组、系统研发组、用户服务组、网络安全组等业务组和共享联盟章程下的多个共建单位,按照“分阶段实施、独立自主创新,边研究、边建设、边服务、边完善”的总体思路共同开展数据资源建设和共享服务。

4.2.2 研制适用于多源、分散地球系统科学数据的标准体系

体系包括:数据共享联盟与平台章程、运行机制与管理条例等10项管理制度,以及元数据标准、数据分类规范、平台开发与共享服务规范等83项标准规范(图1);覆盖数据采集、整合、分析和共享发

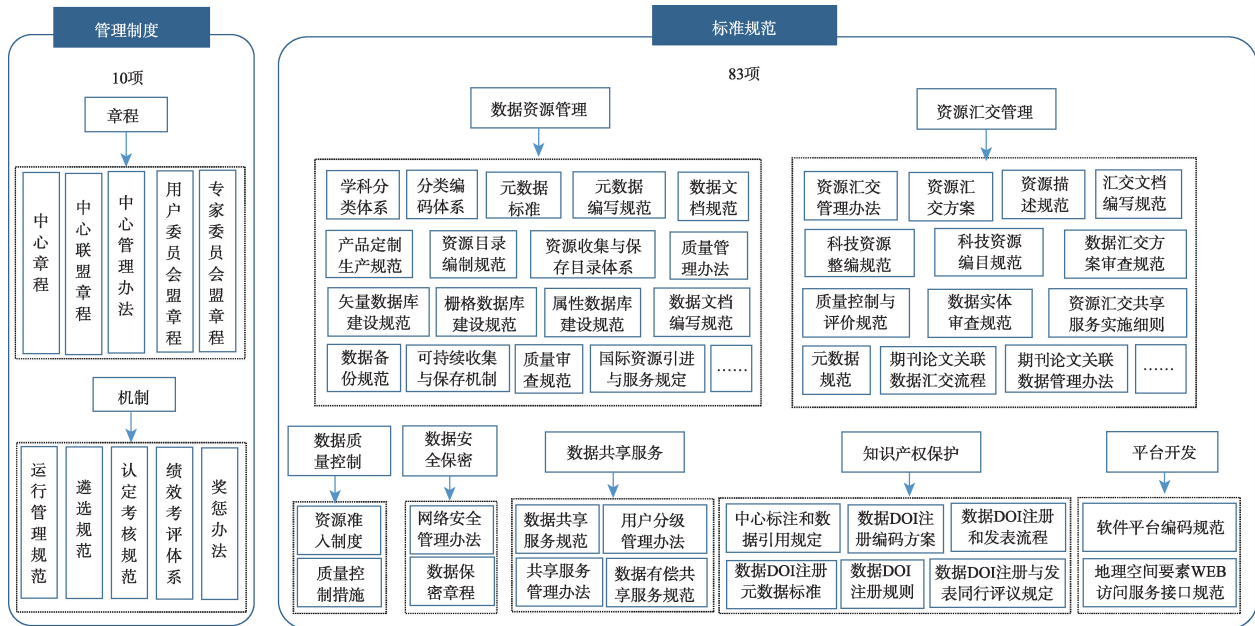


图1 国家地球系统科学数据中心标准体系

Fig. 1 Classification of data sharing standards for the National Earth System Science Data Center

布各个环节。

4.2.3 建成学科面广、多时空尺度的地球系统科学数据库群

整个数据库涵盖5大圈层、58个三级学科,包含1500个国际地学数据站点导航、5个国际数据库镜像、126个专题数据库群,数据资源总量约2.0 PB(图2),包括:联合图书管理部门开展历史资料抢救形成的珍贵的、可二次加工与深度挖掘的科学数据资源,例如涉及地理沿革等重要历史信息的古籍方志和文献,我国历次大型科学考察形成的科学数据、报告、资料及视频,民国地理历史地图集,建国以来全国十大流域的水文资料数据,以及中华人民共和国国家自然、经济、自然区划、土地资源、草地资源、地貌、行政区划变迁、人口与环境变化等多部大型历史地图集;与知名科学家合作研发的具有自主知识产权的数据产品,如全球碳循环关键参数GLASS产品,全球30 m土地覆盖数据产品,MODIS全球500 m植被聚集指数、叶面积指数产品,中国全天候地表温度、土壤湿度、蒸散发产品,全球野火多发区域火灾风险评估数据集,南海岛礁珊瑚共生微生物数据集等;通过国际合作和科学考察获取的国际数据资源,如中蒙古国际经济走廊重点区域社会经济统计年鉴,行政区划、人口、交通、旅游、土地利用、自然资源等历史和现势图集,以及土样、水样、森林样方等调查数据。

4.2.4 自主研发海量、异构数据网络共享技术体系与服务系统

系统框架如图3所示,该系统突破统一数据服务模型数据访问技术和大型网络异常流量预警控制技术,实现了分布式、异构数据资源的透明访问和平台总分一体化的监控管理,保障了共享网络系统的安全运行;系统采用“数据建设-数据审查-数据发布”三级架构保障数据数量,构建数据资源和用户群体分类分级共享方案保证数据资源安全和知识产权,遵照“六位一体”(数据实体、数据样例、数据说明文档、数据缩略图、数据分类和元数据表)标准在线发布数据;系统还嵌入基于时间、空间和主题内容特征的科学数据量化精准语义关联网络实现数据智能推荐,支持用户访问轨迹追踪以针对性地向用户提供主动推送服务以及数据定制服务;此外,系统通过在线云计算平台,结合专业知识图谱体系和专家知识库,实现从数据共享服务到知识服务的转型升级。

4.2.5 拥有持续稳定的国内用户群和不断壮大的国际用户群

据不完全统计,中心实名注册用户达12.5万人,网站访问人次达590万余次,页面访问次数约3.2亿次,累计为8000多个重大科研项目/课题提供了约18 PB的数据支撑和服务,如东北黑土地保护工程、珠港澳大桥、粤港澳大湾区战略规划、地球大



图2 国家地球系统科学数据中心数据资源库群

Fig. 2 The hierarchy of databases from the National Earth System Science Data Center

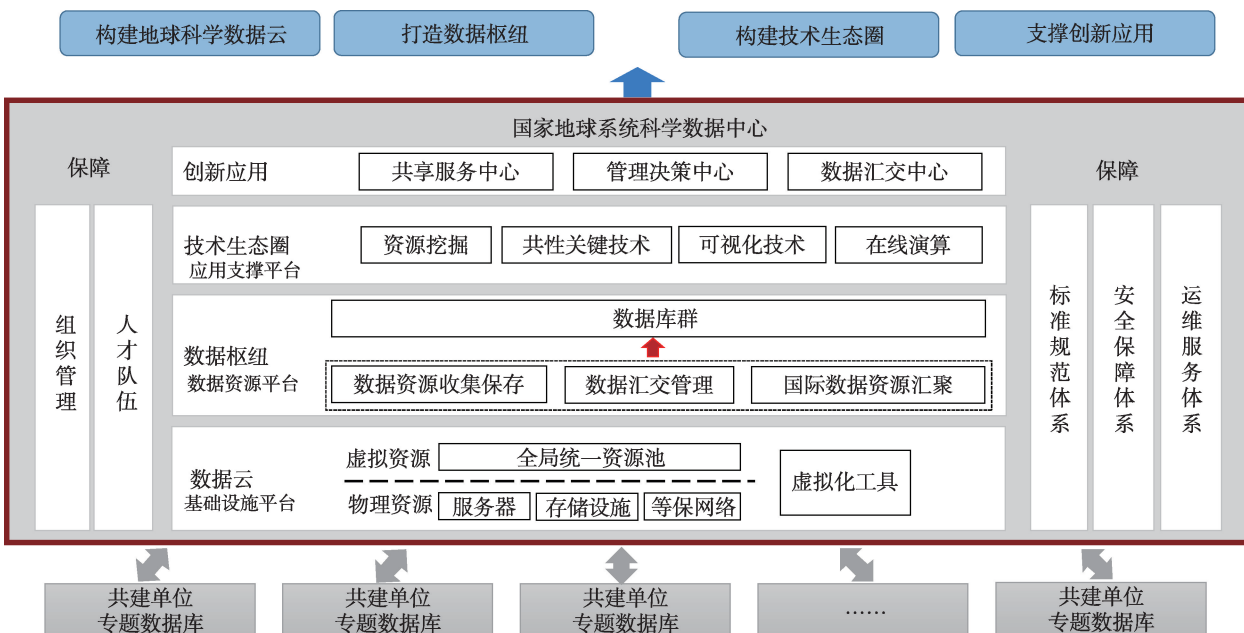


图3 国家地球系统科学数据中心共享服务系统总体框架

Fig. 3 Data sharing service system of the National Earth System Science Data Center

数据科学工程、雄安新区生态评估、国家精准扶贫第三方评估等国家重大项目,青藏高原地质灾害风险评估、孟中印缅区域地缘风险分析、长江三角洲城市群典型受损生态空间修复、丝绸之路非物质文

化景观保护等重点项目。下一步,中心将围绕京津冀协同发展、长江经济带发展、粤港澳大湾区建设、长江三角洲一体化发展和黄河流域生态保护和高质量发展等重大国家战略开展专题数据资源整合

共享,并强化数据分析挖掘能力、提升专业知识服务水平,打造“信息-数据-知识”的一站式服务体系。例如,探讨京津冀联合发展模式及域内资源整合方法,寻找粤港澳多方合作共赢的区域发展模式,探索长江经济带绿色生态发展规律,围绕国家生态文明建设需求服务区域可持续发展,立足2022冬奥会探索冰雪产业为示范的地方经济发展途径等重点方向。

4.2.6 推动数据共享理念的传播与推广

2008年,国家在中心设立首个国家“973”资源环境领域科学数据汇交管理中心,先行探索国家科技项目数据汇交管理模式;2014年,国家以科学数据中心为试点开展国家科技基础性专项项目数据汇交,进一步强化国家科技项目数据汇交与共享应用示范。全国性的数据汇交培训会和数据共享理念宣讲会,强化了科学界对数据共享的认知,科学家及承担单位科学数据汇交意识和能力也得到提升。大批科学家主动将个人科研产出的科学数据贡献到科学数据中心进行共享应用,例如北师大定量遥感团队的GLASS产品,自2018年在中心上线以来陆续发布了叶面积指数(LAI)、反照率(Albedo)、光合有效辐射(PAR)、植被总初级生产力(GPP)等10余种数据产品,在线访问量超10万次、下载超4万次,成为地表能量平衡、地表蒸散发和全球气候变化研究,以及农业、林业、水文、气象等应用领域的重要基础支撑数据。科研项目数据汇交管理实践推动了科研产出科学数据的规范化集成整编与项目成果开放,为国家构建科技项目汇交管理体系、制定科学数据汇交管理规范提供了参考案例。此外,通过参加国际数据共享论坛和学术会议、组织“一带一路”沿线国家数据共享培训会等形式,中心数据共享理念得到国际用户群体广泛认可。当前,数据共享理念在地球系统科学领域正不断深入,数据共享也逐步成为推动地球系统科学发展的关键力量。

5 总结与展望

作为综合性交叉性学科,地球系统科学的进步与发展离不开数据共享,在大数据时代和人工智能时代,数据共享的地位愈发重要。经过多年实践与发展,国家在数据资源管理政策与机制、国家科学数据中心建设、国家科学项目数据汇交、数据出版

与数据产权保护等方面取得了突出成效。在地球系统科学领域,以国家地球系统科学数据中心为核心的实践主体,探索形成符合我国实际的地球系统科学数据资源分类体系,突破数据共享中的元数据处理、分布式互操作、大数据智能分析、数据安全保密等关键技术,构建覆盖数据开放共享全生命周期的目录服务和数据服务体系,在共享机制、标准规范、资源建设、系统研发和数据服务等方面为我国科学数据共享工作提供了宝贵的经验。

当前,地球系统科学领域的数据共享问题仍十分突出,“数据孤岛”现象普遍存在,严重制约了科技创新与进步乃至国家整体科技实力的提高。下一代地球系统科学数据共享模式中,将更加关注数据管理、数据交换、数据共享等全生命周期的技术标准,减少异构系统之间机器操作难度;然而,我国大部分的地球系统科学数据服务机构的标准化程度都比较差,更多采用项目或者部门自己定义、但又与国际规范不对应的技术标准,这无形中也造成了更多的数据孤岛。

地球系统观念的建立和整体上采用多学科交叉集成和多尺度融合应用的研究方法革新使得地球系统科学研究已成为数据密集型的活动。未来,除了科学数据共享外,还需要利用地球大数据,发现、提取和挖掘隐藏在数据背后的规律、知识和新的科学问题,利用现有数据加工生产出新的数据资源,提升数据资源的价值;随着地球系统科学的发展,数据资源类型越来越多、应用场景越来越复杂、数据处理技术越来越强大,公众个性化的需求被激活,将演化出以公众为核心的地球系统科学数据共享流程,定制化的“数据+知识”服务模式有望成为主流趋势。然而,未来观测体系中公众非专业传感器造成的数据异构性,以及各种自然和社会观测中非结构化数据的融入,将对现有数据收集、管理、融合和处理等共享关键技术提出新的挑战。

参考文献(References):

- [1] National Research Council. Earth system sciences: A closer view[M]. Washington, DC: The National Academies Press, 1988.
- [2] Schellnhuber H J. 'Earth system' analysis and the second Copernican revolution[J]. Nature, 1999,402(6761):1579-1583.
- [3] 汪品先.我国的地球系统科学研究向何处去[J].地球科学进展,2003,18(6):21-35. [Wang P X. Earth system sci-

- ence in china [wt3hx]quo vadis? *Advances in Earth Science*, 2003,18(6):837-851.]
- [4] 周秀骥.对地球系统科学的几点认识[J].*地球科学进展*, 2004,19(4):513-515. [Zhou X J. Some cognitions on earth system science[J]. *Advances in Earth Science*, 2004, 19(4):513-515.]
- [5] 诸云强.地球系统科学数据共享关键技术研究[M].北京:科学出版社,2008. [Zhu Y Q. Research on key technologies of earth system science data sharing[M]. Beijing: Science Press, 2008.]
- [6] 孙九林.地球系统科学理论与实践[J].*地理教育*,2006,1(1):4-6. [Sun J L. Theory and practice of earth system science[J]. *Geographical Education*, 2006,1(1):4-6.]
- [7] 黄鼎成,林海,张志强.地球系统科学发展战略研究[M].北京:气象出版社,2005. [Huang D C, Lin H, Zhang Z Q. Study on the development strategy of Earth System Science[M]. Beijing: China Meteorological Press. 2005.]
- [8] 童庆禧.空间对地观测与全球变化的人文因素[J].*地球科学进展*,2005,20(1):1-5. [Tong Q X. Earth observation from space and humman demension for global change studies. *Advances in Earth Science*, 2005,20(1):1-5.]
- [9] 李德仁.脑认知与空间认知——论空间大数据与人工智能的集成[J].*武汉大学学报·信息科学版*,2018,43(12):8-14. [Li D R. Brain cognition and spatial cognition: On integration of geo-spatialbig data and artificial intelligence [J]. *Geomatics and Information Science of Wuhan University*, 2018,43(12):8-14.]
- [10] 郭华东.地球大数据科学工程[J].*中国科学院院刊*,2018,33(8): 818-824. [Guo H D. A project on big earth data science engineering[J]. *Bulletin of the Chinese Academy of Sciences*, 2018,33(8):818-824.]
- [11] Reichstein M, Camps- Valls G, Stevens B, et al. Deep learning and process understanding for data-driven Earth system science[J]. *Nature*, 2019,566(7743):195-204.
- [12] 诸云强,孙九林,廖顺宝,等.地球系统科学数据共享研究与实践[J].*地球信息科学学报*,2010,12(1):1-8. [Zhu Y Q, Sun J L, Liao S B, et al. Earth system scientific data sharing research and practice[J]. *Journal of Geo-Information Science*, 2010,12(1):1-8.]
- [13] 柏永青,杨雅萍,孙九林.国内外科学数据管理办法研究进展[J].*农业大数据学报*,2019,1(3):5-20. [Bai Y Q, Yang Y P, Sun J L. Advances in the study of domestic and foreign scientific data management methods[J]. *Journal of Agricultural Big Data*, 2019,01(3):5-20.]
- [14] 王卷乐,王明明,石蕾,等.科学数据管理态势及其对我国地球科学领域的启示[J].*地球科学进展*,2019,34(3):306-315. [Wang J L, Wang M M, Shi L, et al. The situation of scientific data management and its enlightenment to earth sciences of China[J]. *Advances in Earth Science*, 2019,34(3):306-315.]
- [15] 周小刚,罗云峰.美国国家大气研究中心优先研究领域新特点[J].*地球科学进展*,2006,21(7):751-756. [Zhou X G, Luo Y F. The new characteristics of research priorities of national center for atmospheric research[J]. *Advances in Earth Science*, 2006,21(7):751-756.]
- [16] 张丽丽,温亮明,石蕾,等.国内外科学数据管理与开放共享的最新进展[J].*中国科学院院刊*,2018,33(8):774-782. [Zhang L L, Wen L M, Shi L, et al. Progress in scientific data management and sharing[J]. *Bulletin of the Chinese Academy of Sciences*, 2018,33(8):774-782.]
- [17] 李娟,刘德洪,江洪.国际科学数据共享现状研究[J].*图书馆建设*,2009,32(2):19-25. [Li J, Liu D H, Jiang H. Research on international scientific data sharing[J]. *Library Development*, 2009,32(2):19-25.]
- [18] 黄如花,陈闯.美国政府数据开放共享的合作模式[J].*图书情报工作*,2016,60(19):6-14. [Huang R H, Chen C. Study on the sharing cooperation mode of U. S. open government data[J]. *Library and Information Service*, 2016, 60(19):6-14.]
- [19] Premat C. Can the French Republic be digital? lessons from the last participatory experience on the law-making process[J]. *Policy Analytics, Modelling, and Informatics*, 2018,1(1):247-264.
- [20] 李慧佳,马建玲,王楠,等.国内外科学数据的组织与管理研究进展[J].*图书情报工作*,2013,57(23):130-136. [Li H J, Ma J L, Wang L, et al. Research review on scientific data organization and management at home and abroad[J]. *Library and Information Service*, 2013,57(23):130-136.]
- [21] 欧阳峥峥,青秀玲,顾立平,等.国际数据期刊出版的案例分析及其特征[J].*中国科技期刊研究*,2015,26(5):437-444. [Ouyang Z Z, Qing X L, Gu L P, et al. An study on the case analysis and characteristics of international data journal publishing[J]. *Chinese Journal of Scientific and Technical Periodicals*, 2015,26(5):437-444.]
- [22] 赵蕊菡.科学数据论文的重用现状研究——基于数据期刊“Earth System Science Data”的引文分析[J].*情报理论与实践*,2017,40(11):52-57. [Zhao R H. Research on the status quo of scientific data paper reuse[J]. *Information studies: Theory & Application*, 2017,40(11):52-57.]
- [23] 郎杨琴,孔丽华.科学研究的第四范式吉姆格雷的报告“e-Science:一种科研模式的变革”简介[J].*科研信息化技术与应用*,2010,1(2):92-94. [Lang Y Q, Kong L H. Jim Gray's report of the fourth paradigm of science research: A brief introduction to on e- science: A trans-

- formed scientific method[J]. *e-Science Technology & Application*, 2010,1(2):92-94.]
- [24] 刘灿,王玲,任胜利.数据期刊的发展现状及趋势分析[J].编辑学报,2018,30(4):344-349. [Liu X, Wang L, Ren S L. Landscape and trend analysis of data journals[J]. *Acta Editologica*, 2018,30(4):344-349.]
- [25] 盛小平,武彤.国内外科学数据开放共享研究综述[J].图书情报工作,2019,63(17):6-14. [Sheng X P, Wu T. Review on open sharing of scientific data across the world [J]. *Library and Information Service*, 2019,63(17):6-14.]
- [26] 孙九林,黄鼎成,李晓波.我国科技数据管理和共享服务的新进展[J].世界科技研究与发展,2002,24(5):15-19. [Sun J L, Huang D C, Li X B. The new progress of science data managing and sharing service in China[J]. *World Sci-Tech R & D*, 2002,24(5):15-19.]
- [27] 张洋,肖燕珠.生命周期视角下《科学数据管理办法》解读及其启示 [J]. 图书馆学研究,2019(15):37-43. [Zhang Y, Xiao Y Z. Interpretation and enlightenment of Scientific Data Management Method from the perspective of life cycle[J]. *Researches on Library Science*, 2019 (15):37-43.]
- [28] 张先恩.国家科学数据共享工程[J].科学中国人,2004,12(9):11-13. [Zhang X E. National science data sharing project[J]. *Scientific Chinese*, 2004,12(9):11-13.]
- [29] 中华人民共和国科学技术部.科技部 财政部关于发布国家科技资源共享服务平台优化调整名单的通知(索引号:306-07-2019-031)[Z].2019-06-05.
- [30] 林海,王卷乐.国家重点基础研究发展计划(973)资源环境领域项目数据汇交工作正式启动[J].地球科学进展,2008,23(8):895-896. [Lin H, Wang J L. National key basic research and development plan (973) project data exchange in the field of resources and environment officially launched[J]. *Advances in Earth Sciences*, 2008,23(8): 895-896.]
- [31] 傅伯杰,牛栋,于贵瑞.生态系统观测研究网络在地球系统科学中的作用[J].地理科学进展,2007,26(1):1-16. [Fu B J, Niu D, Yu G R. The roles of ecosystem observation and research network in earth system science[J]. *Advances in Earth Sciences*, 2007,26(1):1-16.]
- [32] 廖顺宝,蒋林.地球系统科学数据分类体系研究[J].地理科学进展,2005,24(6):93-98. [Liao S B, Jiang L. Study on classification system of data for earth system science [J]. *Advances in Earth Sciences*, 2005,24(6):93-98.]
- [33] 王卷乐,林海,冉盈盈,等.面向数据共享的地球系统科学数据分类探讨[J].地球科学进展,2014,29(2):265-274. [Wang J L, Lin H, Ran Y Y, et al. A study of earth system science data classification for data sharing[J]. *Advances in Earth Sciences*, 2014,29(2):265-274.]
- [34] 王卷乐,宋佳,卜坤,等.国家地球系统科学数据共享平台数据分类编目与特征分析[J].中国科技资源导刊,2015,47(6):65-73. [Wang J L, Song J, Bu K, et al. Data classification cataloguing and feature analysis of national data sharing platform of earth system science[J]. *China Science & Technology Resources Review*, 2015,47(6):65-73.]
- [35] Chen Z, Song J, Yang Y P. An approach to measuring semantic relatedness of geographic terminologies using a thesaurus and lexical database sources[J]. *ISPRS International Journal of Geo-Information*, 2018,7(3):98.
- [36] 陈祖刚,杨雅萍.耦合尺度的地理实体空间相关度算法的建立与应用[J].地球信息科学学报,2018,20(1):37-47. [Chen Z G, Yang Y P. A case of establishment and application of spatial correlation degree algorithm for geographic entities coupling scales[J]. *Journal of Geo-Information Science*, 2018,20(1):37-47.]
- [37] 诸云强,刘润达,冯敏,等.分布式地球系统科学数据共享平台研究[J].计算机工程与应用,2009,45(1):245-248. [Zhu Y Q, Liu R D, Feng M, et al. Research on distributed earth system scientific data sharing platform[J]. *Computer Engineering and Applications*, 2009,45(1):245-248.]
- [38] 李善青,郑彦宁,邢晓昭,等.科学数据共享的安全管理问题研究[J].中国科技资源导刊,2019,51(3):1-17. [Li S Q, Zheng Y N, Xing X Z, et al. Study on security managements of scientific data sharing[J]. *China Science & Technology Resources Review*, 2019,51(3):1-17.]
- [39] 郝世博,徐文哲,唐正韵.科学数据共享区块链模型及实现机理研究[J].情报理论与实践,2018,41(11):57-62. [Hao S B, Xu W Z, Tang Z Y. Block chain model of scientific data sharing and its realization mechanism[J]. *Information Studies: Theory & Application*, 2018,41(11): 57-62.]
- [40] 王卷乐.科学数据交换中心元数据关键问题研究:以中国地球系统科学数据共享网为例[D].北京:中国科学院地理科学与资源研究所,2005. [Wang J L. Research on key problems of metadata in scientific data clearinghouse: a case study on China Earth System Science Data Sharing Network[D]. Beijing: Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, 2005.]
- [41] 胡聪.我国科学数据汇交管理现状、问题及对策研究[J].科技创业月刊,2019,32(7):81-84. [Hu C. Research on the current situation, problems and countermeasures of scientific data interchange management in China[J]. *Pioneering With Science & Technology Monthly*, 2019,32(7): 81-84.]

- [42] 陈松景,钱庆,吴思竹,等.精准医学大数据汇交管控模型与应用研究[J].中华医学图书情报杂志,2018,27(10):14-19. [Chen J S, Qian Q, Wu S Z, et al. Big data archiving management and control model of precision medicine and its application[J]. Chinese Journal of Medical Library and Information Science, 2018,27(10):14-19.]
- [43] 王芳,赵洪,马嘉悦,等.数据科学视角下数据溯源研究与实践进展[J].中国图书馆学报,2019,45(5):79-100. [Wang F, Zhao H, Ma J Y, et al. Research and practice progress of data provenance from the perspective of data science[J]. Journal of Library Science in China, 2019,45(5):79-100.]
- [44] 刘润达,赵辉,李大玲.科学数据共享平台之数据联盟模式初探[J].中国基础科学,2010,12(6):27-32. [Liu R D, Zhao H, Li D L. Preliminary study on data alliance model for scientific data sharing platform[J]. China Basic Science, 2010,12(6):27-32.]
- [45] 赵志明.多源异构国土数据整合方法与云共享研究[J].测绘与空间地理信息,2019,42(11):173-176. [Zhao Z M. Research on multi-source heterogeneous land data integration and cloud sharing method[J]. Geomatics & Spatial Information Technology, 2019,42(11):173-176.]
- [46] 杨雅萍,王祎,白燕,等.国家地球系统科学数据中心发展与实践[J].农业大数据学报,2019,1(4):5-13. [Yang Y P, Wang Y, Bai Y, et al. Development and practice of the National Earth System Science Data Center in China[J]. Journal of Agricultural Big Data, 2019,1(4):5-13.]
- [47] Hey T. The fourth Paradigm-Data-Intensive scientific Discovery[M]. Berlin: E-Science and Information Management, 2012.
- [48] 钟华,刘杰,王伟.科学大数据智能分析软件的现状与趋势[J].中国科学院院刊,2018,33(8):812-817. [Zhong H, Liu J, Wang W. Current situation and trend of intelligent analysis software for scientific big data[J]. Bulletin of the Chinese Academy of Sciences, 2018,33(8):812-817.]
- [49] 孙九林,施慧中.中国地球系统科学数据共享服务网的构建[J].中国基础科学,2003,5(1):76-81. [Sun J L, Shi H Z. Construction of earth system science data sharing service network in China[J]. China Basic Science, 2003,5(1):76-81.]