

中国湖泊科学数据共享的路径和方法

马荣华^{1,2*}, 许金朵^{1,2}, 侯渲^{1,2}, 王贞^{1,2}

1. 中国科学院南京地理与湖泊研究所, 南京 210008

2. 国家地球系统科学数据中心湖泊-流域分中心, 南京 210008

摘要: 科学数据重用可以促进科学研究过程的共享与合作, 使数据资源得到更优化的配置。湖泊科学数据作为重要基础性战略资源之一, 支撑国家科技创新、经济社会发展和国家生态文明建设。湖泊科学数据共享逐渐成为我国国家科技创新体系建设的重要内容。本文归纳了湖泊数据生产的主要渠道, 详细介绍了湖泊数据主要数据类型及数据格式, 总结了国内外湖泊科学数据的共享现状, 提出了以规范制度为保障、数据分级分类为基础、服务平台为依托、分级共享服务为手段的一种湖泊科学数据共享的路径和方法, 总结了湖泊数据共享的制度措施, 对湖泊科学数据共享在新形势下的进一步发展提出了建议, 以推动构建湖泊科学数据共享的新模式和湖泊科学研究的新范式, 促进湖泊科学的发展和进步。

关键词: 湖泊数据; 数据共享; 分级分类; 共享路径; 共享方法; 新范式



文献 CSTR:

32001.14.11-6035.csd.2023.0110.zh



文献 DOI:

10.11922/11-6035.csd.2023.0110.zh

文献分类: 地球科学

收稿日期: 2023-05-31

开放同评: 2023-07-12

录用日期: 2023-12-01

发表日期: 2023-12-25

引言

科学数据是科研人员从事基础应用研究过程中产生的数据, 或者通过野外调查、野外试验、野外观测、室内实验、检测检验等方式取得的、用于科学研究的原始数据及其衍生数据, 包括实验记录手稿、电子表格或文本、电子图片或视频、软件代码等。科学数据是科研人员的生命线。不同的科研活动, 会产生不同类型、不同数量的科研数据。科研数据有时候是巨大的, 称为海量数据, 后来又称为大数据 (mega data)。

大数据的概念起源于美国, 由思科、甲骨文、IBM 等公司发起并发展。全球顶级期刊《Nature》在 2008 年 9 月推出了“Big Data”专刊, 其主旨就是为了研讨大数据发展对生物医学、环境科学、互联网技术以及超级计算等多个科技领域所带来的挑战^[1]。2009 年开始, “大数据”成为互联网信息技术行业的热门词汇, 用于指代超出了常规机器或者软硬件工具捕捉、管理和处理能力, 而在一定时间范围内难以处理的数据集合^[2]。2011 年 2 月, 《Science》推出“Dealing with Data”专刊^[3], 通过社会调查的方式, 第一次综合分析了大数据对人们生活造成的影响, 详细描述了人类面临的“数据困境”。2012 年开始, 大数据成为全球科学研究的新热点。大数据具有如下 5V 特征^[4]: (1) 体量大 (Volume), 即从 GB、TB 级跃升到 PB 乃至 EB 级别; (2) 类型多 (Variety), 包括结构化数据、半结构化数据和非结构化数据, 如我们日常生活中所产生的视频监控数据、流媒体数据、射频识别 RFID 数据、感应数据等^[5], 超过 80%都是典型的非结构化数据; (3) 价值大 (Value), 单条数据并无太多价值, 但由巨量的数据汇聚所形成的数据集中蕴含

* 论文通信作者

马荣华: rhma@niglas.ac.cn

着巨大的价值^[6]；(4) 变化快 (Velocity)，这是大数据区别于传统数据挖掘最显著的特征；物联网分分钟都在采集数据，微博内容时时刻刻都在更新，处理速度达到每小时 10 TB 或更高；(5) 准确性好 (Veracity)，数据准确性好、可信度高，即数据质量高。因此，对于大数据的处理和分析更加艰巨、更加复杂，大数据很难由单台计算机处理，须采用分布式计算架构，依赖云计算的分布式处理、分布式数据库、云存储和/或虚拟化技术^[7]。

进入大数据时代，国家高度重视数据共享工作。为进一步加强和规范科学数据管理，以保障科学数据安全，提高开放共享水平，更好支撑国家科技创新、经济社会发展和国家安全^[8]，国务院于 2018 年 3 月印发施行了我国首部针对科学数据管理与共享的法律规范《科学数据管理办法》(国办发〔2018〕17 号)，明确提出了“分级管理、安全可控、充分利用”的原则；2019 年 2 月，中国科学院在落实国家《科学数据管理办法》的基础上，印发了《中国科学院科学数据管理与开放共享办法(试行)》(科发办字〔2019〕11 号)^[9]；2019 年 6 月，科技部和有关部门在前期工作的基础上，组建了 20 个国家科学数据中心；2022 年 12 月 19 日，《中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见》(即“数据二十条”)^[10]对外公布。

湖泊是由湖盆、湖水和水中所含物质以及水生生物组成的自然综合体，是组成地球表层系统的重要单元之一，是包括大气圈、生物圈、岩石圈、陆地水圈等在内的地球表层各圈层相互作用的联结点，是流域生物地球化学要素循环的中心枢纽和物质能量的汇，是流域生态环境变化的指示器^[11]；湖泊生态系统对全球变化的响应显著。湖泊科学数据作为支撑国家科技创新和经济社会发展的重要基础性战略资源之一，湖泊科学数据采集能力也随着我国科技创新能力和投入不断增强而持续提升，具有大数据 5V 的显著特征，大数据为湖泊科学研究提供了一种新途径、一个新思路。湖泊科学数据共享逐渐成为我国国家科技创新体系建设的重要内容，也是大数据时代科技创新和经济社会发展的重要基础。

1 湖泊数据生产渠道

科学数据作为科技创新的基石，湖泊科学数据驱动湖泊科学研究的创新，而湖泊科学数据主要来源于野外考察调查，数据获取费时费力，长时间序列的观测数据获取主要依赖于固定的野外观测站点，特别是进入国家序列的野外科学研究观测台站(太湖湖泊生态系统国家野外科学观测研究站(“太湖站”)、东湖湖泊生态系统国家野外科学观测研究站(“东湖站”)、梁子湖湖泊生态系统国家野外科学观测研究站(“梁子湖站”)、洞庭湖湖泊湿地生态系统国家野外科学观测研究站(“洞庭湖站”)、兴凯湖湖泊湿地生态系统国家野外科学观测研究站(“兴凯湖站”)、鄱阳湖湖泊湿地综合研究站(“鄱阳湖站”))，能够稳定、持续开展监测任务，并将监测数据进行了不同程度的开放共享。其中太湖站、东湖站、梁子湖站通过国家生态系统观测研究网络科技资源服务系统(<http://www.cnern.org.cn/>)共享了 2004–2016 年的常规站点监测数据，共享的数据类型包括了气象、湖泊浮游植物、湖泊底栖动物、湖泊大型水生植物、湖泊微生物、湖泊浮游植物叶绿素、湖泊浮游植物初级生产力、湖泊水物理、湖泊水化学、湖泊沉积物等数据资源。共享方式以协议共享为主，即需要用户提出申请野外台站进行审批同意后数据方可下载使用，部分气象数据可直接下载使用。洞庭湖站、兴凯湖站以及鄱阳湖站通过国家生态科学数据中心(<http://www.nesdc.org.cn/>)发布了洞庭湖 2012–2016 年洲滩土壤元素含量、南荻生态系统通量和常规气象数据，兴凯湖 2011–2018 年水体物理和化学指标监测数据以及“浮毯”沼泽植物组成动态数据，鄱阳湖 2013–2018 年水环境监测数

据以及水体富营养化指标空间分布数据，共享方式为直接下载。

目前，我国上述 6 个国家级野外观测台站的数据开放程度最高，但数据共享的时效也不同程度地存在滞后。据不完全统计，其他我国水域野外科学观测站点还有：抚仙湖高原深水湖泊研究站（中国科学院南京地理与湖泊研究所）、呼伦湖生态系统定位观测研究站（中国科学院南京地理与湖泊研究所）、天目湖流域生态观测研究站（中国科学院南京地理与湖泊研究所）、博斯腾湖湿地生态系统观测点（中国科学院南京地理与湖泊研究所）、青海湖高原湖泊湿地生态环境科学观测野外台站（中国科学院水生生物研究所）、南水北调东线湖泊生态系统湖北省野外科学观测研究站（中国科学院水生生物研究所）、三峡水库香溪河生态系统实验站（中国科学院水生生物研究所）、查干湖生态系统定位观测研究站（中国科学院东北地理与农业生态研究所）、洪湖湿地生态系统湖北省野外科学观测研究站（中国科学院测量与地球物理研究所）、红枫湖湖泊生态系统研究站（中国科学院地球化学研究所）、滇池湖泊生态系统云南省野外观测研究站（昆明市滇池高原湖泊研究院）、内蒙古乌梁素海湿地生态系统国家定位观测研究站（内蒙古农业大学）、华北浅水湖泊湿地生态系统野外科学观测研究站（河北大学）、丹江口湿地生态系统野外科学观测研究站（中国科学院武汉植物园）等。这些台站所积累的科学观测数据暂未对外提供公开开放共享。

除了野外科学研究观测台站外，湖泊科学数据的积累还来源于大量的科研项目的观测数据、调查数据、实验数据以及计算模拟分析数据等^[12]，科研项目类型包括国家重点研发计划、国家科技基础资源调查专项、国家重大专项“水体污染控制与治理科技重大专项（水专项）”、中国科学院先导专项、国家重点基础研究发展计划（973 计划）、国家自然科学基金等项目。

由于受到人力、物力、财力、天气和水文条件的限制，常规人工监测难以进行长时间追踪和监测，只能了解监测断面上的水质或水环境状况。尽管这些数据具有局部和典型样点的代表意义，但对于整个水体而言，数据的代表性并不充分，无法满足实时、大尺度的监测评价需求^[13]。遥感技术具有大范围、快速实时、周期性的特点，弥补了传统人工观测手段的不足，被广泛应用于湖泊水体边界、藻类水华、水生植被以及水质卫星遥感监测产品的生产^[14]。目前基于卫星遥感手段生产典型的、广泛应用的数据集/库主要有：（1）覆盖全球 140 万个表面积至少为 10 公顷湖泊或水库矢量边界的 HydroLAKES 数据库^[15]；（2）1984–2019 年间全球 340 万个面积 $\geq 0.03 \text{ km}^2$ 的全球湖泊边界数据集 GLAKES^[16]；（3）全球面积大于 0.1 km^2 的 248 243 个湖泊（面积总量占全球湖泊面积的 57.1%）的藻华爆发频率及爆发面积数据集^[17]；（4）全球 10 km^2 以上约 14 700 个自然湖泊 2003–2020 年水位-水量变化的空间分布数据集^[18]；（5）全国 1 km^2 以上 2660 个湖泊近四十年（1986–2020）长时序的透明度数据集^[19]。

除此以外，国际上对观测数据的管理与共享，建立了许多部门和行业数据中心来为公众提供数据共享^[20]。如世界数据中心（World Data Center, WDC）、地球观测组织（Group on Earth Observations, GEO）、全球陆地观测系统（Earth observation system, GTOS）、地球观测数据网（The Canadian Earth Observation Network, GeoNET）、世界温室气体数据中心（World Data Centre for Greenhouse Gases, WDCGG）、德国遥感数据中心（the German Remote Sensing Data Center, DFD），旨在促进与推动国际科学数据的共享。

2 湖泊数据类型及格式

湖泊科学是一门研究地球表层系统重要单元的交叉学科^[21]，湖泊科学数据覆盖的学科范围广，

包括了自然地理、水物理、水化学、水文学、气象学、沉积学、环境科学、生态学等多个学科方向。数据类型及数据格式复杂多样，既有观测类表格文本型数据、空间矢量栅格数据，也包含了特定观测仪器产生的特定格式，数据格式涵盖了 xls、txt、doc、shp、tif、hdf、img 等（表 1）。

表 1 湖泊数据类型

Table 1 Types of lake data

序号	学科分类	数据内容	数据类型	数据格式
1	湖泊自然地理	湖泊空间位置、形状、大小、环湖水系、地形、地貌、行政区划、流域界限	矢量/栅格	shp、tif、img
2	湖泊水物理	湖泊 pH、透明度、水温、电导率、含沙量、冰期、盐度等	属性	xls、txt
3	湖泊水化学	湖泊总氮、总磷、氨氮、叶绿素 a、高锰酸盐指数、化学需氧量、硝态氮、亚硝态氮、总碳、无机碳、碱度、铅、汞、砷、镉、铬、镍、铜、铁、锰、锌、钾、钙、钠、镁、碳酸氢根、碳酸根、硫酸盐、氯化物等	属性	xls、txt
4	湖泊水文	湖泊水位、水量、流量等	属性	xls、txt
5	湖泊气象	降雨、蒸发、气温、气压、风速、风向、湿度等	属性	xls、txt
6	湖泊沉积	常规理化性质（含水量、中值粒径、粘土、有机碳、总氮、总磷、钾、钙、钠、镁、铁、铝）、重金属（汞、砷、镉、铅、铜、锌、铬）、测年（ ²¹⁰ Pb、 ¹³⁷ Cs、 ²²⁶ Ra）、有机污染物多环芳烃、无机污染物有机氯农药、硅藻、摇蚊等	属性	xls、txt
7	湖泊生物	湖泊浮游植物、大型水生植物、浮游动物、底栖动物、微生物等	属性/矢量/栅格	xls、txt、shp、tif、img
8	湖泊生态	水色、富营养化、藻华等	属性/矢量/栅格	xls、txt、shp、tif、img

面对多源异构、复杂多样的湖泊数据，其整编需要建立一套标准的规范体系。为提高野外调查数据的规范性，保障监测数据质量，使调查数据具有可比性，编著了《湖泊调查技术规程》^[22]（2015），对湖泊调查、样品采集、分析测试方法进行了标准规范的统一，包括湖泊水质调查规程、湖泊水量调查规程、湖泊生物资源调查规程、湖泊沉积物量调查规程、湖泊沉积物质量调查规程、湖泊卫星遥感调查技术规程以及湖泊数据整编技术规程。在湖泊数据的整编方面，对湖泊编码的规范、湖泊数据加工处理的规范，以及湖泊水质、水量、生物资源、沉积物、卫星遥感监测等各类数据的集成、加工、整编、入库进行了统一的规定。

中国科学院南京地理与湖泊研究所是我国唯一以湖泊-流域为研究对象的国家级科研机构，通过对湖泊科学数据进行了系统的加工整理，构建了湖泊科学数据分类体系，以此为基础建设了中国湖泊科学数据库（<http://www.lakesci.csdb.cn/>）、湖泊-流域科学数据共享平台（<http://lake.geodata.cn/>）。

3 湖泊科学数据的共享现状

目前,在全球科学家的共同努力下,湖泊科学数据的开放共享已初具规模:(1)世界湖泊网:世界湖泊网(<http://www.worldlakes.org/>)^[23]是1998年建立的美国非营利组织,共享世界8000多个湖泊的位置、海拔、深度、水量等信息;(2)WDS(World Data System)成员PANGAEA:PANGAEA^[24]是世界数据中心WDS的成员,收集、发布、共享和出版地球和环境领域科学数据,共收录了800个湖泊和河流相关数据集;(3)中国湖泊科学数据库(<http://www.lakesci.csdb.cn>)^[25]:由中国科学院南京地理与湖泊研究所建设,发布了全国1 km²以上2693个湖泊空间位置、大小、形状以及湖泊相关的水系、流域信息、典型湖泊的水质观测、蓝藻水华监测数据;(4)湖泊-流域科学数据共享平台(<http://lake.geodata.cn>)^[26]:国家地球系统科学数据中心的湖泊-流域分中心,面向湖泊科学领域收集、整编、发布和共享数据,目前收集国内湖泊科学领域约60%数据,发布数据集891个,数据资源量累计30 TB;(5)中国科学院南京地理与湖泊研究所科学数据中心:中国科学院南京地理与湖泊研究所科学数据中心(<http://data.niglas.ac.cn>)^[27]是中国科学院科学数据中心,收集研究所科研项目和论文关联数据,目前已发布数据集568个,数据资源量约10 TB。另外,国家生态系统观测网络(<http://cnern.ac.cn/index.action>)也发布了部分湖泊观测数据。

4 湖泊数据共享的路径和方法

为进一步加强科学数据管理,保障科学数据安全,提高科学数据开放共享水平^[8],2018年,科技部和财政部共同研究制定了《国家科技资源共享服务平台管理办法》^[28],随后国务院办公厅印发了《科学数据管理办法》^[8](国办发[2018]17号)。为落实上述两个《办法》,科技部与财政部以“完善科技资源共享服务体系,推动科技资源向社会开放共享”为建设目标,于2019年在原有的国家科技资源共享服务平台基础上进行优化调整,建设了包含高能物理、基因组、对地观测、地球系统、人口健康、农业、林业、气象、海洋等在内的20个国家科学数据中心^[29]。中国科学院为落实国家《科学数据管理办法》,为了进一步加强中国科学院科学数据管理,保障科学数据安全,提高科学数据开放共享水平,于2019年2月印发了《中国科学院科学数据管理与开放共享办法(试行)》^[9],提出了科研项目数据汇交的要求,加强科研论文关联数据汇交管理,明确科学数据开放共享的原则和主体责任,并随之开展了中国科学院科学数据中心体系建设,由1个总中心、18个学科中心和12个所级中心三类组成,实现中国科学院科学数据的常态化、规范化汇聚管理与安全开放共享。2021年,又印发了《中国科学院科学数据中心管理暂行办法》(科发办字[2021]61号),并发布《中国科学院科学数据工作要点》^[30]。国家相关部门和中国科学院对科学数据管理与共享空前重视。

为保障国家科学数据安全,防止数据流失,把我国的关键科学数据“牢牢掌握在手中”,国家层面针对重大项目制定了系列数据汇交管理办法^[31]。早在2008年,科技部就颁布了《国家重点基础研究发展计划资源环境领域项目数据汇交暂行办法》,启动了“973”计划资源环境领域项目数据汇交工作,实现了已结题的103项“973”计划资源环境领域项目数据汇交清单发布^[32],其中包含湖泊科学相关的项目6项,共汇交数据集285个^[33]。2014年5月,科技部发布了《科技基础性工作专项项目科学数据汇交管理办法(试行)》^[34];截至2019年底,已完成1999–2015年结题验收的科技基础性工作专项项目的数据汇交,共计331项,占有立项项目的76.1%,对外发布共享的湖泊科学相关的科技基础工作专项项目6项^[35],共汇交64个数据集。2017年5月,科技部印发《国家重点研发计

划管理暂行办法》^[36]。2019年12月，科技部办公厅印发《科技计划项目科学数据汇交工作方案（试行）》（国科办基〔2019〕104号），明确了科学数据汇交原则、关联主体与职责、主要内容与流程，加快推进科技计划项目科学数据汇交。中国科学院为进一步优化和完善中国科学院战略性先导科技专项的管理，促进重大成果产出，2022年9月，中国科学院在国家相关法规、办法的基础上，结合先导专项实际情况，印发了《中国科学院战略性先导科技专项科学数据汇交管理实施细则（试行）》（科发规字〔2022〕47号）^[37]，指导先导专项有效、保质地完成数据汇交工作，使中国科学院科学数据作为国家战略性资源得以安全保存和利用。

为加强和规范科学数据管理，保障科学数据的安全，防止数据资源流失，中国科学院还将目光投向了期刊论文关联数据。2021年，中国科学院传播局印发《中国科学院科学传播局关于进一步加强和规范论文关联数据管理的通知》（传播字〔2021〕2号）；2022年印发《中国科学院科学传播局关于从严监管论文关联数据汇交工作的通知》（传播字〔2022〕3号），从严监管论文关联数据汇交工作，开展院内论文关联数据的汇交工作。此项举措保障了院内科研论文关联数据的汇交共享。

在湖泊科学数据的共享路径与方法上，中国科学院南京地理与湖泊研究所湖泊—流域科学数据中心（以下简称“数据中心”），致力于湖泊科学数据的收集、集成、整编、加工整理与共享。为促进湖泊科学数据的共享，节约项目经费使用，中国科学院南京地理与湖泊研究所制定了研究所内数据购买报销制度。截至目前，此项措施收集积累卫星遥感影像、土地利用、社会经济、水文气象、水质等数据量共81GB。购买的数据面向所内人员免费开放共享，此项举措节约了科研人员的大量科研经费。此外，数据中心制定了所内自主设立课题数据汇交制度，鼓励研究所自主设立课题的项目数据汇交。数据中心制定了系列的规章制度来保障中心的可持续运行，在共享服务方面，采取数据分级分类共享策略。

数据中心以湖泊—流域科学数据共享平台为依托，整合的数据资源按照数据类型、共享权限等划分等级，依据一定的共享策略提供共享服务，采取相应的服务模式保证服务质量，使共享资源的服务长期稳定运行（图1）。

（1）共享权限分级

湖泊科学共享权限分级形式包括数据密级分级共享、用户积分共享以及数据提供者审核共享三种形式。

①按照数据密级分级共享：划分为秘密数据、次密数据和公开数据。

②按照用户积分共享（主要按照数据的贡献大小来分类）：将数据集按照不同专题类型或区域进行划分，分别赋予不同分值，注册用户需具有等同于或大于该数据集的分值才可以下载数据。用户通过平台汇交数据，数据经过审核接收后即可获取积分。按照用户积分高低划分用户共享权限。

③数据提供者审核共享：数据提供者向平台汇交数据，数据申请者在数据共享使用时，需要提出申请，数据提供者同意后，平台可向申请者发放数据，提供共享服务。

（2）服务流程

湖泊科学数据共享的服务流程有以下几种模式：①离线服务共享；②在线服务的数据密级和积分策略共享；③在线服务的数据提供者审批共享。

①离线服务共享：用户离线填写数据使用申请，通过离线的形式邮寄到数据中心，由数据中心服务人员受理，根据用户申请的数据范围和类型来判断分发数据的方式。如果申请的数据是公共类型的数据，则采用离线的方式，通过邮寄等分发给用户；如果该数据的使用需要获取数据拥有者的同意，则由平台服务人员联系数据拥有者，咨询数据拥有者意见，征得同意后，采用离线的方式和

用户签署数据保密协议，平台服务人员收到数据申请者签署后的保密协议后，再将数据离线分发给用户；如果数据属于密级数据，则和用户离线签署数据保密协议，按照国家密级数据使用的相关规定，指导用户使用。



图 1 我国湖泊科学数据共享路径和方法

Figure 1 Shared strategies and methodologies of lake science data

②数据密级和积分策略共享：数据用户提出数据使用申请，数据中心收到使用申请后，根据数据的可公开性进行审核，公开共享数据可通过国家地球系统科学数据中心湖泊-流域分中心 (<http://lake.geodata.cn>) 或中国科学院南京地理与湖泊研究所科学数据中心 (<http://lake.data.ac.cn>) 进行申请获取；若是非公开数据，数据中心判断根据用户积分情况进行审核。

③数据提供者审批共享：对该类共享条件的数据，数据中心将湖泊科学数据的审核权限转交至数据提供者，根据数据提供者的审核意见来提供数据。

通过这 3 种共享服务方式，保障了密级数据资源的安全、数据提供者的知识产权以及处理权，使数据申请者最大程度获益，同时也激励数据申请者转变为数据提供者，扩大数据资源积累。

(3) 质量控制

湖泊科学数据共享数据，遵守“FAIR”原则，即可发现 (Findable)、可访问 (Accessible)、可互操作 (Interoperable) 和可重复 (Reproducible)。共享数据的质量，采取一体化流程进行质控，贯穿整个数据生命周期。原始数据知识产权归属数据提供者，原始数据质量由数据提供者负责。数据中心在数据发布共享的过程中，数据进行二次校验。第一次由数据中心对数据进行质量校验，发现异常值和数据生产者进行确认；第二次发布数据前邀请业内专家对数据质量把关，对存在问题的数据反馈给数据生产者进行修改校正。共享发布的过程中，数据质量采取“五位一体”的原则，即每个

发布的数据集要具有数据说明文档、数据缩略图、元数据、数据样例，以及数据实体。

(4) 共享成效

数据中心自 2011 年至 2023 年 9 月以来，通过上述的共享方式与共享流程，面向国内外用户共提供了 12 941 次共享服务（图 2），开展典型服务 82 项，专题服务 92 项；服务各类科研项目 3640 项，其中国家重大项目 1035 项，国家自然科学基金 1343 项，省部级项目与国际合作项目 1262 项；支撑发表 SCI/EI 等各类论文共 1788 篇，支撑授权软著 49 项，发明专利 37 项，共享成效显著。

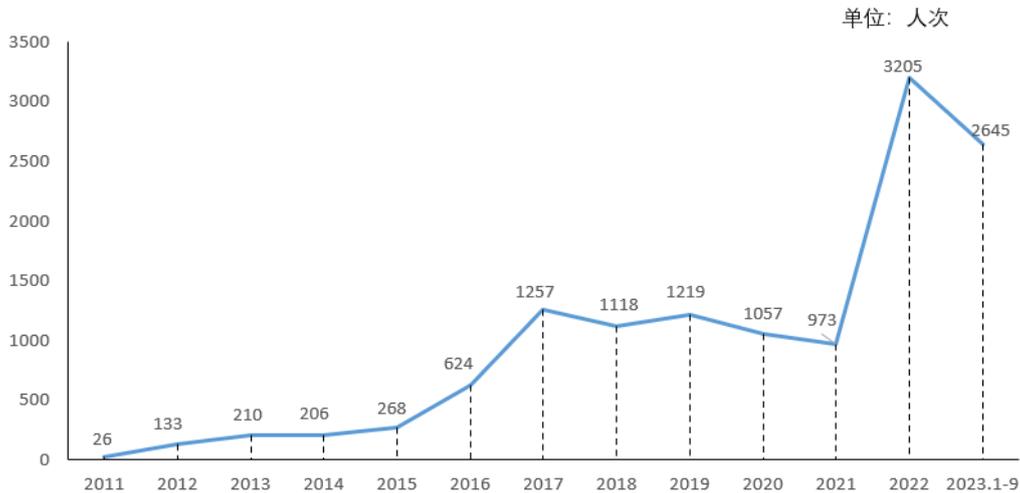


图 2 数据中心 2011–2023 年日常服务情况

Figure 2 Daily shared-service of data center from 2011 to 2023

5 中国湖泊科学数据共享保障

国务院办公厅在 2018 年 3 月 17 日印发了《科学数据管理办法》，其目的就是为了进一步规范科学数据管理，提高科学数据共享的水平^[8]。对促进科学数据共享工作具有开创性意义；2019 年 2 月，中国科学院在此基础上印发了《中国科学院科学数据管理与开放共享办法（试行）》，为科学数据开放共享提供规范制度的同时，也对科学数据安全进行保护，加强科学数据管理，是中国科学院实施国家大数据战略的重要举措^[9]。

中国科学院南京地理与湖泊研究所为保障湖泊科学数据的共享与使用，制定了系列的规章制度，包括《湖泊数据整编规程》《中国湖泊沉积物底质调查数据整编技术规程》《中国湖泊卫星遥感调查技术规程》《湖泊–流域科学数据元数据规范 V1.0》《湖泊–流域科学数据整编技术规范 V1.0》《湖泊–流域科学数据质量控制规范（试行稿）》和《湖泊–流域科学数据整合集成方案》。制定了一系列数据管理与数据共享的规范制度：《中国科学院南京地理与湖泊研究所科学数据管理办法（试行）》《湖泊–流域科学数据共享条例》《湖泊–流域数据共享平台数据共享积分政策（试行）》《湖泊–流域科学数据保密协议》《湖泊–流域科学数据转让保密协议》《中国科学院南京地理与湖泊研究所科研项目数据汇交管理办法》和《中国科学院南京地理与湖泊研究所期刊论文数据汇交方案》。

6 湖泊科学数据共享存在的问题

中国湖泊科学数据资源数据共享服务虽取得了一定的成效，但是仍然存在一些问题和困难：①

国内从事湖泊科学研究机构众多，湖泊科学数据资源分散，不同单位（大学、科研院所、政府事业单位）、不同部门（院系、处室）、不同科研团队、不同台站，信息孤岛以及数据壁垒问题，难以短时间内消除；②有价值的核心科学数据资源大多掌握在科研人员手中，这些数据是科研人员的核心竞争力，导致共享意愿不强烈；③在馈赠数据提供者方面，目前主要以论文标注、致谢等方式，导致了推广及传播效果不是很显著。

7 展望

随着大数据时代的到来，信息化已进入了新阶段，数据密集型科研已经成为大数据时代科学发展的新模式，科学数据开放共享已成科技界的普遍共识^[38]。党中央、国务院高度重视数字经济发展，“推动大数据技术产业创新发展、构建以数据为关键要素的数字经济、运用大数据提升国家治理现代化水平、运用大数据促进保障和改善民生、切实保障国家数据安全”的战略，是我国构筑大数据时代国家综合竞争新优势的必由之路。未来很长的一段时间内，湖泊科学数据共享要在下面几个方面持续发展：

（1）成立湖泊数据共享联盟，建立同盟机制，最大限度地聚合湖泊学科数据资源，建立长时间序列的、要素信息完整的数据集；确立湖泊科学数据资产地位、定量价值地位，打破行业、部门之间的数据壁垒，助力湖泊数据融入应用场景，释放湖泊数据资源真正价值；

（2）构建集成由数据资源、开源平台、API、数据基础设施、数据挖掘与分析、数据应用等板块构成的湖泊大数据生态系统；持续构建可伸缩的湖泊生态环境治理保护的计算体系结构，提高湖泊数据处理和再分析能力；建立完善的湖泊数据安全机制，实现湖泊数据要素安全流通，释放科学数据资源价值。

（3）国家层面，大力推进数据共享事业的发展，形成数据共享的浓厚氛围和利剑行动，推进人大立法，制定实施《中华人民共和国数据共享法》。

致 谢

本研究得到中国科学院“十四五”网络安全和信息化专项（CAS-WX2022SDC-SJ05，CAS-WX2021SF-0306）资助。

作者分工职责

马荣华（1972—），男，山东临沂人，研究员，研究方向湖泊环境遥感与数据共享。主要承担工作：论文框架构建、主体撰写以及文章校验。

许金朵（1982—），女，江苏徐州人，高级工程师，研究方向数据库建设、数据共享。主要承担工作：文章修改与校对。

侯渲（1998—），男，甘肃天水人，工程师，研究方向数据库建设、数据共享。主要承担工作：负责相关规章制度的整理与校对。

王贞（1983—），女，河南安阳人，工程师，研究方向湖泊水文。主要承担工作：文章的修改与校对。

参考文献

- [1] Nature. Big Data[EB/OL]. (2008-09-03)[2023-05-27]. <http://www.nature.com/news/specials/bigdata/index.html>.
- [2] 李国杰, 程学旗. 大数据研究: 未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J]. 中国科学院院刊, 2012, 27(6): 647-657. DOI: 10.3969/j.issn.1000-3045.2012.06.001. [Li G J, Cheng X Q. Research Status and Scientific thinking of Big Data[J]. Bulletin of Chinese Academy of Sciences, 2012, 27(6): 647-657. DOI: 10.3969/j.issn.1000-3045.2012.06.001.]
- [3] Science. Special online collection: Dealing with data[EB/OL]. (2011-02-11)[2023-05-27]. <https://www.science.org/toc/science/331/6018>.
- [4] 李清泉, 李德仁. 大数据 GIS [J]. 武汉大学学报(信息科学版), 2014, 39(6): 641-644. DOI: 10.13203/j.whugis20140150. [LI Q Q, LI D R. Big Data GIS[J]. Geomatics and Information Science of Wuhan University, 2014, 39(6): 641-644. DOI: 10.13203/j.whugis20140150.]
- [5] 刘泽照, 朱正威. 大数据平台下的社会稳定风险评估: 研究前瞻与应用挑战[J]. 华东理工大学学报(社会科学版), 2015, 30(1): 78-85. DOI: 10.3969/j.issn.1008-7672.2015.01.010. [LIU Z Z, ZHU Z W. Research Prospects and Application Challenges of Social Stability: Risk Assessment in the Context of Big Data[J]. Journal of East China University of Science and Technology (Social Science Edition), 2015, 30(1): 78-85. DOI: 10.3969/j.issn.1008-7672.2015.01.010.]
- [6] 樊伟红, 李晨晖, 张兴旺, 等. 图书馆需要怎样的“大数据”[J]. 图书馆杂志, 2012, 31(11): 63-68+77. DOI: 10.13663/j.cnki.lj.2012.11.017. [FAN WH, LI C H, ZHANG X W, et al. How Do Libraries Need “Big Data”?[J]. Library Journal, 2012, 31(11): 63-68+77. DOI: 10.13663/j.cnki.lj.2012.11.017.]
- [7] 宋关福, 钟耳顺, 李绍俊, 等. 大数据时代的 GIS 软件技术发展[J]. 测绘地理信息, 2018, 43(01): 1-7. DOI: 10.14188/j.2095-6045.2017478. [SONG G F, ZHONG E S, LI S J, et al. Development of GIS Software Technology in the Era of Big Data[J]. Journal of Geomatics, 2018, 43(01): 1-7. DOI: 10.14188/j.2095-6045.2017478.]
- [8] 关于印发科学数据管理办法的通知[EB/OL]. (2018-03-17)[2023-05-27]. https://most.gov.cn/xxgk/xinxifenlei/fdzdgnr/fgzc/gfxwj/gfxwj2018/201804/t20180404_139023.html [Notice of General Office of the State Council on publishing the Measures for the Management of Scientific Data[EB/OL]. (2018-03-17)[2023-05-27]. https://most.gov.cn/xxgk/xinxifenlei/fdzdgnr/fgzc/gfxwj/gfxwj2018/201804/t20180404_139023.html]
- [9] 中国科学院科学数据管理与开放共享办法(试行)[EB/OL]. (2019-02-11). [2023-05-27]. <https://www.cas.cn/gzzd/ggsw/zh/202108/P020210827414998424076.pdf>. [Measures for the management of scientific data and open sharing of Chinese Academy of Sciences (Trial) [EB/OL]. (2019-02-11). [2023-05-27]. <https://www.cas.cn/gzzd/ggsw/zh/202108/P020210827414998424076.pdf>.]
- [10] 中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见[EB/OL]. (2022-12-19). [2023-05-27]. http://www.gov.cn/zhengce/2022-12/19/content_5732695.htm. [Opinions of the State Council on Data Elements Playing a More Important Role in Building a Data Infrastructure System[EB/OL]. (2022-12-19) [2023-05-27]. http://www.gov.cn/zhengce/2022-12/19/content_5732695.htm.]

- [11] 李世杰, 窦鸿身, 舒金华,等. 我国湖泊水环境问题与水生态系统修复的探讨[J]. 中国水利, 2006, (13): 14-17.[LI S J, DOU H S, SHU J H, et al. Water environmental problems of lakes and water ecosystem recovery in China[J]. China Water Resources, 2006, (13): 14-17.]
- [12] 刘杨, 陈晓玲, 全志薇,等. 国家级科学数据中心建设对吉林省科学数据中心建设的示范效应[J]. 中国管理信息化, 2022, 25(03): 173-178. DOI: 10.3969/j.issn.1673-0194.2022.03.050. [LIU Y, CHEN X L, QUAN Z W, et al. Demonstration Effect of National Scientific Data Center Construction on Scientific Data Center Construction in Jilin Province [J]. China Management Informationization, 2022, 25(03): 173-178. DOI: 10.3969/j.issn.1673-0194.2022.03.050.]
- [13] 马荣华, 段洪涛, 唐军武. 湖泊水环境遥感[M]. 北京: 科学出版社, 2010. [MA R H, DUAN H T, TANG J W. Remote Sensing Of Lake Water Environment[M].Beijing: Science Press, 2010.]
- [14] 洪勇豪, 亓郑男, 张丽丽. 遥感大数据在水利中的应用及发展[J]. 水利信息化, 2019(3): 25-31. DOI: 10.19364/j.1674-9405.2019.03.005. [HONG Y H, QI Z N, ZHANG L L. Application and Prospect of large remote sensing data in hydrology[J]. Water Resources Informatization, 2019(3): 25-31. DOI: 10.19364/j.1674-9405.2019.03.005.]
- [15] MESSAGER M L, LEHNER B, GRILL G, et al. Estimating the volume and age of water stored in global lakes using a geo-statistical approach[J]. Nature Communications, 2016, 7: 13603. DOI: 10.1038/ncomms13603.
- [16] PI X H, LUO Q Q, FENG L, et al. Mapping global lake dynamics reveals the emerging roles of small lakes[J]. Nature Communications, 2022, 13(1):5777. DOI:10.1038/s41467-022-33239-3.
- [17] HOU X J, FENG L A, DAI Y H, et al. Global mapping reveals increase in lacustrine algal blooms over the past decade[J]. Nature Geoscience, 2022, 15(2):130-134. DOI: 10.1038/s41561-021-00887-x.
- [18] LUO S X, SONG C Q, KE L H, et al. Satellite laser altimetry reveals a net water mass gain in global lakes with spatial heterogeneity in the early 21st century[J]. Geophysical Research Letters, 2022, 49(3): e2021GL096676. DOI:10.1029/2021GL096676.
- [19] ZHANG Y B, ZHANG Y L, SHI K, et al. Remote sensing estimation of water clarity for various lakes in China[J]. Water Research, 2021, 192:116844. DOI:0.1016/j.watres.2021.116844.
- [20] [1] 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(01):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(01):1-8.]
- [21] 王苏民, 窦鸿身. 中国湖泊志[M]. 北京: 科学出版社, 1998. [WANG S M, DOU H S. China Lake Records[M]. Beijing: Science Press, 1998.]
- [22] 中国科学院南京地理与湖泊研究所. 湖泊调查技术规程[M]. 北京: 科学出版社, 2015. [Nanjing Institute of Geography & Limnology Chinese Academy of Sciences. Technical regulations for lake survey [M]. Beijing: Science Press, 2015.]
- [23] LakeNet. World Lakes Website[EB/OL]. (2014-03-22)[2023-05-27].<http://www.worldlakes.org/>.
- [24] World Data System. PANGAEA[EB/OL]. (2023-05-27)[2023-05-27].<https://www.pangaea.de/>.
- [25] 中国科学院南京地理与湖泊研究所. 中国湖泊科学数据库 [EB/OL]. (2016-10-11)[2023-05-27].<http://www.lakesci.csdb.cn>. [Nanjing Institute of Geography & Limnology Chinese Academy of Sciences. China Lake Database[EB/OL]. (2016-10-11)[2023-05-27]. <http://www.lakesci.csdb.cn>.]

- [26] 国家地球系统科学数据中心.湖泊-流域科学数据共享平台[EB/OL]. (2020-12-01)[2023-05-27].<http://lake.geodata.cn>. [National Earth System Science Data Center. Lake and Watershed Scientific Data Sharing Platform[EB/OL]. (2020-12-01)[2023-05-27].<http://lake.geodata.cn>.]
- [27] 中国科学院南京地理与湖泊研究所.中国科学院南京地理与湖泊研究所科学数据中心[EB/OL]. (2023-05-27)[2023-05-27].<http://data.niglas.ac.cn>. [Nanjing Institute of Geography & Limnology Chinese Academy of Sciences. Scientific Data Center of Nanjing Institute of Geography & Limnology Chinese Academy of Sciences[EB/OL]. (2023-05-27)[2023-05-27].<http://data.niglas.ac.cn>.]
- [28] 关于印发《国家科技资源共享服务平台管理办法》的通知[EB/OL]. (2018-02-13)[2023-05-27].https://www.most.gov.cn/xxgk/xinxifenlei/fdzdgnr/fgzc/gfxwj/gfxwj2018/201802/t20180224_138207.html [Notice on the Issuance of the Management Measures of the National Science and Technology Resources Sharing Service Platform[EB/OL]. (2018-02-13)[2023-05-27].https://www.most.gov.cn/xxgk/xinxifenlei/fdzdgnr/fgzc/gfxwj/gfxwj2018/201802/t20180224_138207.html]
- [29] 国家科技资源共享服务平台优化调整名单 [EB/OL]. (2019-06-11)[2023-05-27] https://www.gov.cn/xinwen/2019-06/11/content_5399105.htm [Optimization and adjustment list of national science and technology resources sharing service platform [EB/OL]. (2019-06-11)[2023-05-27] https://www.gov.cn/xinwen/2019-06/11/content_5399105.htm]
- [30] 关于印发《中国科学院科学数据工作要点》的通知[EB/OL]. (2020-12-24)[2022-11-22]. <http://www.digital.cas.cn/gzwj/ygzd/202101/P020210107504156285987.pdf>. [Notice of publishing Key Points of Scientific Data Work for Chinese Academy of Sciences[EB/OL]. (2020-12-24)[2023-05-27].<http://www.digital.cas.cn/gzwj/ygzd/202101/P020210107504156285987.pdf>.]
- [31] 高雅丽. 在科技管理中, 让科学数据“开放共享”[N]. 中国科学报,2022-06-08(001).DOI:10.28514/n.cnki.nkxsb.2022.001260. [GAO Y L. In Science and Technology Management, Let Scientific Data “Open Sharing” [N]. China Science Daily, 2022-06-08(001).DOI:10.28514/n.cnki.nkxsb.2022.001260.]
- [32] 白燕, 杨雅萍, 王祎. 科技基础性工作专项资源环境领域项目数据汇交进展与分析[J]. 中国科技资源导刊, 2020, 52(5): 52-62. DOI: 10.3772/j.issn.1674-1544.2020.05.008. [BAI Y, YANG Y P, WANG Y. Data Collection Progress and Data Analysis of National Special Program of Basic Research Works for Science and Technology in Resource and Environment Field[J]. China Science & Technology Resources Review, 2020, 52(5): 52-62. DOI: 10.3772/j.issn.1674-1544.2020.05.008.]
- [33] 国家地球系统科学数据中心.国家 973 计划资环领域项目数据项目列表[EB/OL]. (2023-05-27)[2023-05-27]. <https://www.geodata.cn/project973/projects.html?> [National Earth System Science Data Center. Data Projects List in the Resource & Environmental Protection Field of the National 973 Plan[EB/OL]. (2023-05-27)[2023-05-27].<https://www.geodata.cn/project973/projects.html?>]
- [34] 科技基础性工作专项项目科学数据汇交管理办法(试行)[EB/OL]. (2014-05-13)[2023-05-27]. <http://www.most.gov.cn/tztg/201405/W020140521318416401810.doc>. [Management measures of data collection for national special program of basic research works for science and technology(Trial) [EB/OL]. (2014-05-13)[2023-05-27]. <http://www.most.gov.cn/tztg/201405/W020140521318416401810.doc>.]
- [35] 国家地球系统科学数据中心.科技基础性工作专项项目列表[EB/OL]. (2023-05-27) [2023-05-

- 27].<http://www.geodata.cn/project/projects.html?> [National Earth System Science Data Center. List of National Special Program of Basic Research Works for Science and Technology[EB/OL]. . (2023-05-27) [2023-05-27]. <http://www.geodata.cn/project/projects.html?>]
- [36] 关于印发《国家重点研发计划管理暂行办法》的通知[EB/OL]. (2017-06-22)[2023-05-27].https://www.most.gov.cn/xxgk/xinxifenlei/fdzdgknr/fgzc/gfxwj/gfxwj2017/201706/t20170628_133796.html. [Notice of Interim Measures for the Management of National Key Research and Development Programs[EB/OL]. (2017-06-22)[2023-05-27].https://www.most.gov.cn/xxgk/xinxifenlei/fdzdgknr/fgzc/gfxwj/gfxwj2017/201706/t20170628_133796.html.]
- [37] 中国科学院印发战略性先导科技专项科学数据汇交管理实施细则[EB/OL]. (2022-09-01)[2023-05-27].<https://mp.weixin.qq.com/s/WwNbr6re5rEppmLGizxMw>. [The Chinese Academy of Sciences Launched the Implementation Rules for the Management of Scientific Data Collection of Strategic Priority Research Program [EB/OL]. (2022-09-01)[2022-11-22]. <https://mp.weixin.qq.com/s/WwNbr6re5rEppmLGizxMw>.]
- [38] 李云婷, 温亮明, 张丽丽, 等. 科学数据共享系统的现状与趋势[J]. 农业大数据学报, 2019, 1(4):86-97. DOI:10.19788/j.issn.2096-6369.190409. [LI Y T, WEN L M, ZHANG L L, et al. The Status and Trends of Scientific Data Sharing Systems[J]. Journal of Agricultural Big Data, 2019, 1(4):86-97. DOI:10.19788/j.issn.2096-6369.190409.]

论文引用格式

马荣华, 许金朵, 侯渲, 等. 中国湖泊科学数据共享的路径和方法[J/OL]. 中国科学数据, 2023, 8(4). (2023-12-25). DOI: 10.11922/11-6035.csd.2023.0110.zh.

Strategies and methodologies for facilitating lake data sharing in China

MA Ronghua^{1,2*}, XU Jinduo^{1,2}, HOU Xuan^{1,2}, WANG Zhen^{1,2}

1. Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, P.R. China

2. Lake-Watershed Sub Data Center, National Earth System Science Data Center, Nanjing 210008, P.R. China

*Email: rhma@niglas.ac.cn

Abstract: The reuse of scientific data can foster the sharing and collaboration of research processes, facilitating a more efficient allocation of data resources. Lake scientific data, as an important fundamental strategic resources, play a crucial role in driving national scientific and technological innovation, fostering economic and social development, and advancing the construction of a national ecological civilization. The sharing of scientific data pertaining to lakes has gradually emerged as an important part of China's national scientific and technological innovation system construction. This research synthesizes the primary channels for lake data production, offering a comprehensive overview of the main data types and formats of lake data from the perspective of lake discipline as an interdisciplinary discipline, and summarizes the current situation

of sharing scientific data on lakes both domestically and internationally. Furthermore, this paper also proposed strategies and methodologies for facilitating sharing scientific data related to lakes based on standardized systems, data classification, service platforms, and hierarchical sharing services. In this study, we concluded the institutional measures for lake data sharing, and provided insights for the further development of lake science data sharing under new circumstances. This research is undertaken with the aim of fostering the transformation of lake science data sharing modes under the new paradigm of lake science research.

Keywords: lake data; data sharing; hierarchical classification; shared strategy; sharing methodology; new paradigm