

我国科学数据中心建设现状及其运行机制研究

戴力新 戴 琦

(江苏省科技资源统筹服务中心, 江苏南京 210000)

摘要: 完善科学数据共享和管理体系是提升国家创新体系整体效能的重要任务之一。在调研并梳理当前我国科学数据中心建设现状的基础上, 分析我国科学数据中心内部运行机制, 从组织管理、资源支撑、数据规范、数据安全、数据共享、产权标示、交流合作 7 个方面提炼出我国科学数据中心运行的基本框架, 为今后我国布局科学数据中心、推动其持续运行发展提供启示与借鉴。

关键词: 科学数据中心; 科学数据; 科学数据中心建设; 运行机制; 数据共享

DOI: 10.3772/j.issn.1674-1544.2023.05.001

CSTR: 15994.14.issn.1674.1544.2023.05.001

中图分类号: G203

文献标识码: A

Research on the Construction Status and Operation Mechanism of China's Scientific Data Center

DAI Lixin, DAI Qi

(Jiangsu Province Science and Technology Resources Coordination Service Center, Nanjing 210000)

Abstract: Improving the scientific data sharing and management system is one of the important tasks to improve the overall efficiency of the national innovation system. By systematically sort out the current situation of the construction of domestic scientific data centers, the paper summarizes and analyzes the internal operation mechanism of domestic scientific data centers. The basic framework for the operation and management of China's scientific data center is proposed from seven aspect: organization and management, resource support, data specification, data security, data sharing, property rights identification, and exchange and cooperation. This paper provides inspiration and reference for the future construction and sustainable operation and development of scientific data centers in China.

Keywords: scientific data center, scientific data, scientific data center construction, operation mechanism, data sharing

0 引言

科学数据是支撑科技创新和经济社会发展的基础性和战略性资源, 对科技进步、经济增长、社会发展、国家安全发挥关键作用。作为数据管理、开放共享的重要载体, 科学数据中心是国家

信息基础设施的重要组成部分, 也是现代科研进入数据驱动“第四范式”的基础保障, 因此加快建设科学数据中心已成为实施科技自立自强战略的现实需求。运行机制是指某一系统事物正常运行时各创新要素进行相互作用的动力、方式及运行原理等^[1], 是建设科学数据中心的核心理念。

作者简介: 戴力新 (1966—), 男, 江苏省科技资源统筹服务中心主任, 副研究员, 主要研究方向为科技资源管理服务、科技成果转化; 戴琦 (1995—), 女, 江苏省科技资源统筹服务中心研究实习员, 主要研究方向为科技资源管理服务 (通信作者)。

基金项目: 江苏省科技创新能力建设计划项目“江苏省科学数据中心建设”(BM2022011)。

收稿时间: 2023 年 4 月 6 日。

研究分析国内科学数据中心现有运行机制,对于进一步完善我国科学数据载体布局建设,推进科学数据共享利用、服务和支撑科技创新,健全国家创新体系以及实施创新驱动发展战略具有重要意义。

近年来,我国愈加重视科学数据管理工作。2018年,国务院办公厅印发《科学数据管理办法》,将科学数据管理上升到国家安全高度。在大型数据管理设施建设方面,我国于2002年启动“国家科学数据共享工程”,在此基础上于2019年建成20个国家科学数据中心,自此我国科学数据共享网络正式形成。多个省市也陆续发布了科学数据管理细则,相继启动相关数据设施建设。强化科学数据载体支撑,完善科技资源共享体系,推动数据要素高效配置,统筹推进国家、区域科学数据中心建设,也是提升国家创新体系整体效能的重要任务之一。鉴于此,本文以国内科学数据中心为研究对象,系统梳理建设现状,分析其内部运行机制,对今后布局更多科学数据中心、促进我国科技创新体系发展具有较强的现实意义。

1 相关研究回顾

科学数据中心是科学数据管理与共享研究的具体实践。作为相对年轻的科技载体,近年来与此相关的研究呈上升趋势。其中,科学数据中心建设现状与运行模式经验研究是当前研究的主要方向。一方面,由于国内科学数据中心建设起步较晚,借鉴国外科学数据中心成熟的发展经验以指导国内科学数据中心的发展是常见的思路。如国家科技基础条件平台中心调研了欧美国家科学数据中心发展情况,为我国科学数据中心体系建设提供了基础经验^[2];吴思竹等^[3]从资源、技术、服务视角提供了美国国家生物技术信息中心和欧洲生物信息学研究所的建设经验与启示。另一方面,随着国内科学数据中心体系的逐步建立,分析科学数据中心建设现状日益深入,进一步提出未来布局的设想。涂志芳等^[4]认为数据中心是当前我国科学数据管理与共享的主要模式,且认为

国家科学数据中心体系、中国科学院科学数据中心体系和高校及图书馆数据管理平台是我国科学数据中心的3种主要类型,并从依托单位、服务内容、服务成效和国际影响力等方面对其管理共享模式进行了分析;刘杨等^[5]总结了10个国家级科学数据中心的组织体系和管理模式,并为吉林省建设省级科学数据中心提出了对策建议;杨远征等^[6]、鲁峰等^[7]对各国家科学数据中心下属分中心的资源建设方案进行了探索;储节旺等^[8]讨论了在长三角地区建设区域科学数据中心的必要性,并提出可能的建设方案。同时,由于科研数据突出的学科差异性,主要针对某领域国家科学数据中心的建设发展及共享服务模式而展开的研究也十分活跃。如范国梅等^[9]、赵国锋等^[10]、张连翀等^[11]、杨雅萍等^[12]分别对微生物、地震、对地观测、地球系统等国家科学数据中心的主要服务、发展现状、未来建设等方面进行了介绍和分析。

整体而言,当前关于国内外科学数据中心建设发展的相关研究存在以下问题:一是对国内科学数据中心建设现状缺乏系统梳理。多侧重于某个领域或某个科学数据中心,缺乏全面的调研与分析;二是对科学数据中心运行机制较少探讨。我国科学数据中心的发展历程较短,相关的机制模式尚不成熟,因此当前大部分学者重点对科学数据中心的建设及管理服务模式经验进行梳理,其研究聚焦于数据技术、数据标准、数据应用等方面,但整体而言缺乏机制性的深入研究,科学数据中心的运行机理仍然没有被完全揭示。因此,本文拟在对各类科学数据中心开展系统性调研的基础上,总结提炼具有推广意义的运行机制,以期进一步推动我国科学数据事业发展,促进创新要素流动共享。

2 建设现状

我国科学数据中心的试点探索最早始于21世纪初。国内科学数据中心建设逐渐形成体系,并根据建设级别将我国科学数据中心分为国家级、省级两类。其中,国家级科学数据中心包括

科技部、财政部和中国科学院分别部署建设的科学数据中心。

2.1 国家科学数据中心

2002年，科技部联合诸多科研机构倡导实施“科学数据共享工程”，并发布了一系列相关办法和标准文件。至2008年后，科学数据共享工作转向建设国家各科技基础条件平台^[13]。2018年出台了《科学数据管理办法》。随后，在此文件精神指导下，国家在2019年对原有基础条件平台开展优化调整工作，研究形成“国家高能物理科学数据中心”等20个国家科学数据中心（表1）^[14]，要求各中心推进相关领域科技资源向国家平台汇聚与整合，强化科技资源开发应用与分析挖掘利用，健全安全保障体系，提升科技资源使用效率和科技创新支撑能力。

2.2 中国科学院国家科学数据中心

2019年，中国科学院启动中国科学院科学数据中心体系部署和建设工作，其体系由总中心、学科中心、所级中心3类组成，是中国科学院科学数据管理与开放共享保障机制的重要组成部分

（表2）^[15]。总中心由中国科学院计算机网络信息中心承担建设，负责为中国科学院科学数据资源提供存储及其他技术保障、建设运行中国科学院科学数据公共服务平台，制定通用标准规范、推动科学数据交叉融合应用；学科中心共18个，主要负责汇交、整编、集成本学科领域的科学数据，促进本学科领域科学数据的深度应用。截至2022年年底，依托中国科学院内单位建设的11家国家科学数据中心均为中国科学院学科中心；所级中心共13个，负责推动本法人单位科研项目数据、论文关联数据等的汇交、整编，并定期将数据备份至学科中心或总中心，在中国科学院内对所级中心建设情况进行年度绩效考核，2019年、2020年分别评选出8家优秀所级中心。

2019年，中国科学院印发了《中国科学院科学数据管理与开放共享办法（试行）》，明确“科技项目数据汇交与管理”以及“论文关联数据的汇交与管理”机制。面向科学数据汇交业务场景，中国科学院编制了数据中心建设、项目科学数据汇交、汇交数据质量管理和共享等方面的

表1 20个国家科学数据中心

序号	国家科学数据中心名称	依托单位	主管部门
1	国家高能物理科学数据中心	中国科学院高能物理研究所	中国科学院
2	国家基因组科学数据中心	中国科学院北京基因组研究所	中国科学院
3	国家微生物科学数据中心	中国科学院微生物研究所	中国科学院
4	国家空间科学数据中心	中国科学院国家空间科学中心	中国科学院
5	国家天文科学数据中心	中国科学院国家天文台	中国科学院
6	国家对地观测科学数据中心	中国科学院遥感与数字地球研究所	中国科学院
7	国家极地科学数据中心	中国极地研究中心	自然资源部
8	国家青藏高原科学数据中心	中国科学院青藏高原研究所	中国科学院
9	国家生态科学数据中心	中国科学院地理科学与资源研究所	中国科学院
10	国家材料腐蚀与防护科学数据中心	北京科技大学	教育部
11	国家冰川冻土沙漠科学数据中心	中国科学院寒区旱区环境与工程研究所	中国科学院
12	国家计量科学数据中心	中国计量科学研究院	国家市场监督管理总局
13	国家地球系统科学数据中心	中国科学院地理科学与资源研究所	中国科学院
14	国家人口健康科学数据中心	中国医学科学院	国家卫生健康委
15	国家基础学科公共科学数据中心	中国科学院计算机网络信息中心	中国科学院
16	国家农业科学数据中心	中国农业科学院农业信息研究所	农业农村部
17	国家林业和草原科学数据中心	中国林业科学研究院资源信息研究所	国家林业和草原局
18	国家气象科学数据中心	国家气象信息中心	中国气象局
19	国家地震科学数据中心	中国地震台网中心	中国地震局
20	国家海洋科学数据中心	国家海洋信息中心	自然资源部

表 2 中国科学院科学数据中心体系

总中心 中国科学院计算机网络信息中心	
学科中心（18 个）	所级中心（13 个）
中国科学院高能物理科学数据中心 中国科学院基因组科学数据中心 中国科学院微生物科学数据中心 中国科学院空间科学数据中心 中国科学院天文科学数据中心 中国科学院对地观测科学数据中心 中国科学院青藏高原科学数据中心 中国科学院生态科学数据中心 中国科学院冰川冻土沙漠科学数据中心 中国科学院地球系统科学数据中心 中国科学院基础学科公共科学数据中心 中国科学院化学化工科学数据中心 中国科学院干细胞与再生医学科学数据中心 中国科学院植物科学数据中心 中国科学院地球大数据科学数据中心 中国科学院海洋科学数据中心 中国科学院凝聚态物质科学数据中心 中国科学院脑科学数据中心	中国科学院昆明植物研究所科学数据中心 中国科学院水生生物研究所科学数据中心 中国科学院东北地理与农业生态研究所科学数据中心 中国科学院南海海洋研究所科学数据中心 中国科学院广州生物医药与健康研究所科学数据中心 中国科学院南京土壤研究所所级中心 中国科学院兰州化学物理研究所所级中心 中国科学院南京地理与湖泊研究所所级中心 中国科学院中国科学技术大学所级中心 中国科学院生态环境研究中心所级中心 中国科学院南京地质古生物研究所所级中心 中国科学院昆明动物研究所所级中心 中国科学院遗传与发育生物学研究所所级中心

一系列工作指南，并由中国科学院科学数据总中心研制了一系列支撑科学数据中心业务的软件工具。启动科学数据中心建设以来，中国科学院已积累超过 80 PB 的科学数据存储量。

2.3 省级科学数据中心

自《科学数据管理办法》发布之后，全国各地积极推动落实。截至 2022 年年底，约 15 个省份发布配套科学数据管理实施细则。在省级科学数据中心建设方面，贵州省、甘肃省、广东省走在前列（表 3）。贵州省依托其大科学装置及配套算力资源，自 2018 年起陆续启动建设了生物学大数据中心、SKA 数据中心等 5 个中心。2020 年年底，甘肃省成立气候变化科学数据中心及自然灾害科学数据中心，同步设立甘肃省科学数据总中心（生态环境科学数据中心）。

广东省以计划项目的方式支持省级科学数据中心建设。自 2021 年起，广东省在平台基地及科技基础条件建设项目中设立“科学数据中心”专题，并给予每项省级财政一次性事前无偿资助经费 300 万元，共建设“1+6+5”个科学数据中心，包含 1 个总中心和 10 个领域中心。总中心

由广东省科技基础条件平台中心承担，包括农业、林业、地理、中医药、基因组、珠江流域水环境、广东海洋气象、临床医学、病原微生物、碳中和等领域，并支持一批国家科学数据中心在粤建立分中心。

3 运行机制分析

《科学数据管理办法》对国内科学数据中心的职责架构进行了明确的规定^[16]，科学数据中心运行的核心是对科学数据的管理和利用。目前，以“科学数据生命周期”为框架建立科学数据管理体系得到了大部分学者的认可。因此，本文参照储节旺等^[17]的研究，将科学数据生命周期分为数据收集、数据处理、数据保存、数据共享、数据分析、数据再利用的做法，将科学数据中心运行机制解析为组织管理、资源支撑、数据规范、数据安全、数据共享、产权标示、交流合作 7 个主要方面，并结合我国科学数据中心实践经验进行分析，以期提炼出科学数据中心运行的基本框架，为今后我国布局更多科学数据中心提供启示与借鉴。

表3 国内省级科学数据中心建设情况

省份	省级科学数据中心名称	依托单位
甘肃省	甘肃省科学数据总中心（甘肃省生态环境科学数据中心）	中国科学院西北生态环境资源研究院
	甘肃省气候变化科学数据中心	兰州大学
	甘肃省自然灾害科学数据中心	甘肃省地震局
贵州省	贵州贵安超级计算中心	贵安新区科创产业发展有限公司
	贵安新区生物医学大数据中心	中国科学院上海生命科学研究院
	遥感（地球）数据中心	贵州省遥感中心、国家遥感中心贵州分部
	天文大装置FAST-SKA科学数据中心（SKA亚洲区域中心）	中国科学院国家天文台（北京）
	科技文献数据中心	贵州师范大学、贵州大学、贵州省科学技术情报研究所
广东省	广东省科学数据服务管理中心	广东省科技基础条件平台中心
	岭南特色农业科学数据中心	广东省农业科学院农业经济与信息研究所
	广东林业科学数据中心	广东省林业科学研究院
	广东省粤港澳大湾区地理科学数据中心	广东省国土资源测绘院
	粤港澳大湾区地理科学数据中心	广东省科学院广州地理研究所
	广东省中医药科学数据中心	广州中医药大学第二附属医院
	广东省基因组科学数据中心	深圳华大生命科学研究院
	珠江流域水环境科学数据平台	生态环境部华南环境科学研究所
	广东省海洋气象科学数据中心	中国气象局广州热带海洋气象研究所
	广东省临床医学科学数据中心	广东省卫生健康委员会事务中心
	广东省高致病性病原微生物科学数据中心	中山大学
	广东省陆地—海洋生态系统碳中和科学数据中心	中山大学

3.1 组织管理机制：协同联动，数据加速流动

组织管理机制主要指科学数据中心的外部组织形态和内部组织体系，是明确统一目标，组织各部门、各层次、各单位建立起稳定的相互关系，具体表现为科学数据中心中纵向的等级分工及横向的协同关系架构^[18]。我国科学数据中心带有明显的体制内色彩，目前均呈现出“政府引导，多方联动”的顶层组织架构。这种架构不仅有利于政府部门、科研机构等多条线通力协作，还能串联数据来源到共享端的一系列创新要素，构建跨越国家、省、市等多个层级的运行网络。

国内科学数据中心由具体领域优势高校院所牵头，组建领域内数据共建共享联盟，形成“1总中心+N分中心”的管理服务体系（图1），由于学科性质、建设环境等差异，分中心一般分为4种，即数据站点、领域分中心、区域分中心、应用推广中心。①数据站点作为科学数据中心的资源末端，长期渗透在数据采集一线，为数据中心输送原始基础数据，对于海洋、极地、气象等对实地实时数据有较高要求的数据中心来说，数

据站点是其数据网络的底层支撑；②领域分中心依托各细分领域优势科研单位建设，主要服务于数据类目体系搭建及重点科研项目，是数据存储、分级分类、清洗加工、分析挖掘等环节的重要主体，是数据网络的核心驱动；③区域分中心多依托省级平台建设，是科学数据应用示范基地，能够实现国家和地方创新资源互流，解决服务产业共性难题和满足企业个性需求，推动地方创新与经济发展，是数据网络的配置枢纽；④应用推广中心往往附属于总中心，致力于探索科学数据新的应用场景与服务模式。整体而言，牵头单位与共建单位依据各自的资源优势，按照不同的数据资源内容和数据来源进行分工，执行统一的数据管理制度和数据质控标准，保证数据资源质量及其共享服务水平。

3.2 资源支撑机制：叠加融合，数据多点集散

资源支撑机制主要为科学数据中心获得持久、稳定、多维度、高质量的科学数据资源提供保障。由于科学数据内涵广泛，存在于科技创新研究的各个环节，因此设定全面完善的数据供应

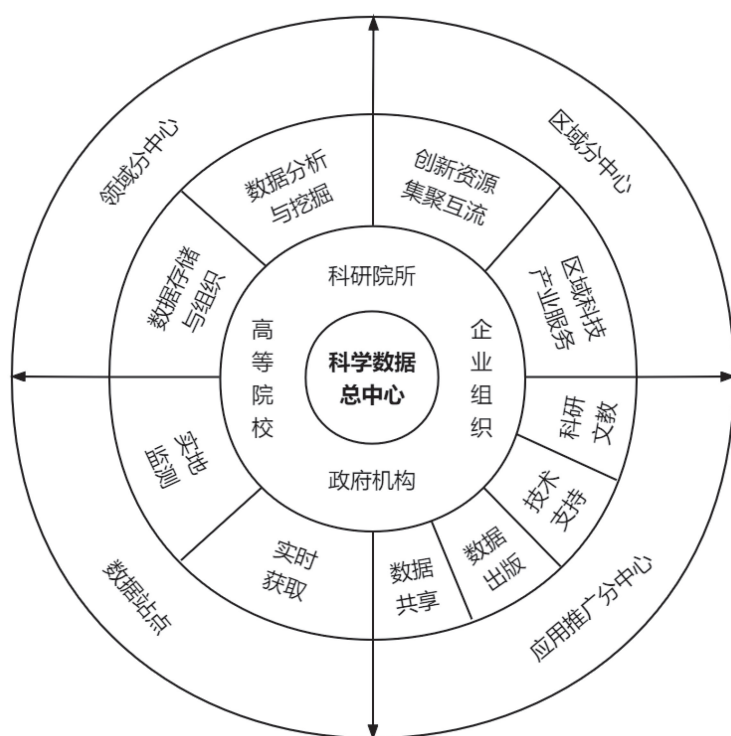


图1 我国科学数据中心“1+N”管理服务体系示意

链、形成科学合理的数据分工架构显得尤为重要。目前，国内科学数据中心主流的资源支撑机制为“内建+外汇”模式，本着“凸显领域、叠加优势、多线汇流”的原则，各科学数据中心形成了独特的资源群落结构。

依据各自的领域定位，数据来源主体覆盖本中心依托建设单位、领域内其他高校院所、相关科技计划项目承担单位、专业数据库商、一线科研人员等，已基本形成形态广覆盖、范围多子域、结构多层次的数据格局。从数据形态看，包括观测试验数据、论文数据、项目汇交数据等，根据各科学数据中心特质与领域属性，不同形态数据占比不一，如国家天文科学数据中心以天文观测数据资源为主，国家材料腐蚀与防护科学数据中心则以检测试验数据为基础；从数据子域看，对于大多数科学数据中心而言，其领域边界渐渐弱化，除了其所属主要学科分支子域外，还吸纳了交叉学科、多尺度的综合集成数据；从数据层级看，原始数据、中间数据、产品数据和衍

生数据都是科学数据中心资源整合的范围，活跃在一线的科研单位及数据采集站提供基础的第一手科学数据资料，通过数据清洗、挖掘分析细化数据粒度，提升科学数据资源质量与规模，借此延伸出更多的专业数据服务。

3.3 数据规范机制：建立标准，数据一致有效

科学数据的规范化程度是数据发现的基础，也是衡量科学数据中心科学数据资源水平的重要指标。由于科学数据具有多域性、异构性、敏感性等诸多特质，要整体提升和加强科学数据的规范化水平，实现数据之间的协调一致、层次分明、可互操作，必须对数据标准体系进行顶层设计和系统构建^[19]。

在推进科学数据中心建设的过程中，学科或行业的领域性特质往往也会催生新的数据标准需求。因此，要关注科学数据中心的领域定位，充分考虑科学数据管理服务的共性和差异，健全各自全生命周期的科学数据标准体系^[19]。广义而言，科学数据标准主要分为通用标准和专用标准

两类^[20]。通用标准用于整体规范数据采集、描述、汇交、安全、服务、评价等全生命周期管理标准，专用标准主要规定不同领域的数据管理与共享要求^[17]。国家基础学科公共科学数据中心牵头编制的通用标准较多，目前已形成采集与整理、元数据、系统与接口、数据管理、数据质量、数据服务等多种类型的标准规范 40 余项，相关成果对各领域科学数据的应用、开放和共享具有重要意义；专用标准主要由各领域国家科学数据中心牵头或参与编制，如地球系统科学数据中心建立了包括 18 项标准规范在内的共享标准规范体系，分属机制条例、数据管理、平台开发、用户服务 4 类，且各标准与地球系统科学数据共享网的子网整合，实现了全网应用^[21]。

3.4 数据安全机制：分级分类，数据治理有序

数据安全保障机制是指通过必要的技术和管理措施，保护科学数据在其全生命周期中免受破坏性外力和非授权操作的侵害，保证科学数据的机密性、完整性和可用性^[22]。在各科学数据中心的现有实践中，数据的分级分类被公认为是实现科学数据全生命周期安全管理的重要手段。

数据的分级分类是指按照一定的方法和原则，划分数据的类别与级别，从而评估数据的安全风险等级，并基于此确定数据相应的安全管理措施。数据类别通常按照数据的领域、来源等属性进行划分，数据级别大多基于其敏感性、风险性等维度进行设定，前者常常作为后者的基础和参考^[23]。根据数据分级分类的结果，针对科学数据的采集、存储、加工、分析、共享等全生命周期环节，可以采用不同的安全管理措施与技术手段进行干预，如数据备份、数据脱敏、权限访问等。

科学数据的安全管理要求与其领域属性密切相关。目前，生命健康、人类基因、地理测绘等关键领域是数据安全研究的重点，相关科学数据中心也已在内部制定了科学数据安全管理制度、数据分级分类框架及标准，并配备相应的安全管理团队及安全保障技术工具，在制度和执行层尽可能地避免数据安全隐患。如国家计量科学数据

中心在《中华人民共和国数据安全法》正式实施以前就制定了比较详细的数据分级分类方案，并在 2020 年发布的《国家计量科学数据中心数据分级分类管理办法》中，根据数据的内容和形式，按照层次分类法将计量科学数据分为两级并对其进行编码^[24]，规范有序地促进科学数据开放共享。

3.5 数据共享机制：多元链路，服务延展升级

开放共享是建设科学数据中心载体的核心目标，数据共享机制是激发各主体开展跨区域、跨行业、跨部门间的数据共享所建立的制度、方法、模式等。根据科学数据资源安全属性、本体价值及用户需求，科学数据中心制定不同的共享策略，逐渐形成相对统一的共享体系，通过设定数据共享的分流机制打开多元化服务链路，增强数据延展性，推动数据共享向知识共享服务升级。

科学数据的安全属性与共享属性密切交织，从安全管理的角度，科学数据共享可分为“无条件、有条件、暂不对外”3 类。针对无条件共享型数据，科学数据中心依托线上平台构建资源目录，实时更新本中心生产、汇交、研制的各类数据，并专题推送热门、重点数据集供用户下载；针对有条件共享型数据，科学数据中心背靠国内一流的科研单位及专业研发团队，围绕国家重大战略性或用户专项需求，制定个性化服务方案，整合算力等基础设施，开展数据资源深度处理加工，为用户提供数据存储、数据监测、数据分析挖掘、数据出版、技术培训等专业化服务和产品支持；针对当前存在的不能满足用户需求、开发利用程度不高、暂时无法对外共享数据的问题，科学数据中心举办征集活动、设立开放课题等方式广泛吸纳优质数据资源，强化重点数据战略研究与利用。

从应用面来看，科学数据主要应用于科学研究和产业创新 2 个方向，分别对应于“基础数据共享、数据工具服务”2 类模式。在科研型应用场景中，高校院所的科研人员常常针对某项目课题开展系统化的研究，由于具备较强的分析处理

与甄别能力,基础或专题数据库一般可以满足此类用户需求;在产业型应用场景中,用户群主要为研发型企业,在产品创新的过程中往往需要大量的计算与模拟仿真,因此智能便捷的可视化软件工具更能直接服务于此类用户,在提供数据工具服务过程中,服务提供方也能与企业进行深入对接,创造更多产学研合作的契机。

3.6 产权标示机制:鉴定出版,数据确权流转

科学数据权益不清是影响共享的重要因素之一。数据产权标示机制通过确定数据权利主体及权利内容来解决数据的权属问题,从而保障数据生产者合法权益,提升数据信度和效度,推动数据安全可靠流转与利用。数据出版通过同行评议等质量控制手段,使数据达到可引用和追溯的状态^[25],被认为是一种有效的数据确权方式。目前,数据出版主要有学术论文关联出版、数据仓储平台出版、数据论文出版等模式,而相关国家科学数据中心已有阶段性的探索成果。

在数据仓储层面,中国科学院计算机网络信息中心于2015年发布具有国际化服务能力的论文关联数据存储库平台,即科学数据银行(Science DB)。科学数据银行联合学术期刊,构建“论文+关联数据”的新型出版生态,同时鼓励科研人员在论文投稿前将数据上传到数据银行,配以数据唯一标识,国家基础学科公共科学数据中心依托科学数据银行为用户线上提供数据集出版及论文出版两种服务。其他类似的数据仓储还包括国家空间科学数据中心建设的空间科学论文数据仓储(SADR)、国家生态科学数据中心建设的国家生态科学数据存储库(EcoDB)等。

在期刊层面,中国科学院计算机网络信息中心于2016年创办了《中国科学数据(中英文网络版)》期刊,开启了科学数据出版时代。目前专刊收录空间、天文、计量学等多个学科领域相关数据论文超过500篇。随后,中国科学院地理科学与资源研究所等联合创办实体数据与论文关联出版期刊《全球变化数据学报》,中国农业科学院农业信息研究所创办我国农业领域首个大数

据专业期刊《农业大数据学报》,都在不同程度上促进了本学科科学数据的出版与引用。

3.7 交流合作机制:开放互动,数据国际通用

长期以来,诸多学科领域的权威数据中心分布美国、欧洲、日本等发达国家和地区。我国科学数据中心的建立,既是为了保障国家科学数据主权,也是为了在有序的框架下推动数据共享和国际交流。为产生更多高质量科学数据成果,科学数据中心也需积极开展国际交流与合作。

国家基因组科学数据中心自2019年起就对标国际核酸序列共享联盟(INSDC),并邀请这个联盟专家担任国际顾问。其不仅通过建设新冠病毒信息库为国内外疫情防控作出巨大贡献,更成功加入了INSDC成为合作伙伴,并建立了以我国为主的国际生物多样性和健康大数据共享联盟(BHDB),当前已与12个国家的28个机构建立了数据共享和科研合作关系。目前,这个中心已连续5年被国际刊物《核酸研究》评价为与美国国立生物技术信息中心(NCBI)、欧洲生物信息学研究所(EBI)并列的全球主要生物数据中心。从国际“散兵”到“伙伴”再到“核心”,国家基因组科学数据中心真正打出了国家品牌,让数据在多边流通中增值^[26]。

其他科学数据中心也对标国际,着力强化数据的可信度和认可度。国家空间科学数据中心是世界数据系统(WDS)的正式成员,2019年成为被美国地球物理学会(AGU)认可的学术论文关联数据仓储库,其与国家人口健康科学数据中心数据仓储(PHDA)均通过了Core Trust Seal全球核心可信存储库国际认证^[27];2020年,国家青藏高原科学数据中心通过Nature数据期刊Scientific Data认证,成为Nature及其子刊文章投稿时可靠和便捷的数据仓储中心;2021年,国家生态科学数据中心数据存储库(EcoDB)正式被国际科研数据仓储目录re3data.org和FAIRsharing收录,成为AGU认定的数据存储库,并被开放存取存储库目录OpenDOAR收录。这些国际通行证为我国在国际舞台上赢得了话语权,无疑也将把我国的数据共享事业推向一个新的高度。

4 我国科学数据中心可持续发展运行的建议

组织管理、资源支撑、数据规范、数据安全、数据共享、产权标示、交流合作作为科学数据中心运行机制的7个主要方面，为今后我国布局更多科学数据中心提供了可资参考的基本框架。科学数据中心的建设和运行是一个长期工作，为保证其未来发展的可持续性，本文提出以下建议。

4.1 明确领域布局，加强顶层设计

科学数据中心应首先明确其非营利性质的科技资源共享服务平台定位，并围绕此定位选定其目标区域及领域。目标区域的选定取决于其建设级别，国家级平台紧密衔接重大科技战略，侧重于为基础研究和源头创新、保障国家安全、提升国家科技核心竞争力提供基础支撑，地方级平台则聚焦于区域主导产业，主要服务于地方科技创新活动。在目标领域的选取方面，一要本着“避免重复”的原则，与现有相关科学数据中心的领域布局区分开。二要结合国家重大战略或地方优势学科和产业特色，在前沿领域或优势产业上下功夫，充分发挥数据资源对于重点科技创新活动的条件保障和支撑作用。在定位明晰的基础上，科学数据中心需确定依托建设单位，加强与部门、单位间协作，建立起“1+N”的组织运行体系，构建集科学数据资源、数据政策、数据规范、数据服务等于一体的全链条制度体系。

4.2 引导多方共建，放大集聚效应

从科学数据中心数据资源支撑及共享机制中不难看出，多方参与、协同推进是扩大资源汇聚规模、提升资源服务效能的关键。一方面，要充分发挥领域内共建共享联盟合力，集聚多源多类多样科学数据资源。联合领域内高校院所等科研单位实现各类科学数据资源的横向整合，联动各级主管部门推动相关科技计划项目科学数据汇交，并加强与国内其他科学数据中心对接，实现优质数据资源的纵向交换。另一方面，要充分运用市场化手段引导高校院所、专业机构积极参与，探索“数据开源”“数据发榜”的共享机制，

利用金融、技术、数据等多重要素激励，推动多方协作与服务模式创新，加速科学数据融合与挖掘利用，合力打造具有本领域或本区域特色的科学数据服务品牌。

4.3 拓展转化途径，提升共享效率

在我国大数据产业快速起步，数据要素市场蓬勃发展的背景下，我们理应探索科学数据在数字经济时代新的表达式，而成果转化正是实现数据资源高效配置、提升数据流通利用效率的新思路。通过挖掘分析，科学数据资源可以形成可发现、可验证、可复用的科学数据产品，科研人员通过智慧劳动对此产品赋予了经济价值，但只有待数据确权 and 交易等规则确定后，科学数据才可以等同于专利、技术等传统型科技成果，进行数据转化。数据成果的确权与转化将成为时代的呼唤，但落到实践还存在一系列的挑战。对于科学数据中心来说，唯一能做的就是在做好科学数据分级分类安全管理的前提下，强化数据多源集聚和融合分析，推动形成一批长序列、多时相、细粒度、高质量的数据集产品，研究科学数据评价机制，探索培育科学数据成果的新路径。

5 结语

科学数据中心作为数据管理的重要载体和科学数据汇集管理、开放共享的重要基础设施，是大数据时代科技创新能力提升的重要保障。本文通过梳理分析国内科学数据中心建设现状，总结其建设运行机制，以期为我国今后布局更多科学数据中心提供参考。未来科学数据中心建设还需充分借鉴国内现有模式经验，在顶层体系设计、多方共建机制、数据转化探索等方面重点发力，努力为建设科技强国提供强有力的数据支撑与保障。

参考文献

- [1] 宋姗姗,王金平,邱科达.协同创新视域下我国智库创新运行机制研究[J].中国高校科技,2021(8):40-44.
- [2] 石蕾,高孟绪,徐波,等.欧美建设发展科学数据中心

- 的经验及对我国的启示[J].中国科技资源导刊, 2022, 54(3): 31-36.
- [3] 吴思竹, 王安然, 修晓蕾, 等. 欧美生物医学科学数据中心建设及启示[J].数字图书馆论坛, 2022(4): 2-10.
- [4] 涂志芳, 杨志萍. 我国科学数据管理与共享实践进展: 聚焦两种主要模式[J].图书情报知识, 2021(1): 103-112.
- [5] 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(3): 173-178.
- [6] 杨远征, 徐超, 唐世林, 等. 南海及邻近海区科学数据中心资源建设初探[J].热带海洋学报, 2021, 40(6): 14-22.
- [7] 鲁峰, 王立华, 徐硕. 渔业科学数据中心建设研究[J].农业大数据学报, 2019(3): 57-70.
- [8] 储节旺, 杨婷婷. 长三角跨区域科学数据中心建设研究[J].数字图书馆论坛, 2022(6): 58-64.
- [9] 范国梅, 孙清岚, 史文聿, 等. 国家微生物科学数据中心数据资源服务与应用[J].微生物学报, 2021, 61(12): 3761-3773.
- [10] 赵国峰, 庞丽娜, 李丽. 国家地震科学数据中心的数据资源与共享服务[C]//第三届地球物理信息前沿技术研讨会论文摘要集.[出版地不详]: [出版者不详], 2021: 86-88.
- [11] 张连翀, 李国庆, 李静. 国家对地观测科学数据中心建设及服务[J].卫星应用, 2021(7): 8-14.
- [12] 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(4): 5-13.
- [13] 阮冰颖, 刘桂锋, 苏文成. 我国科学数据管理实践探索的回顾与展望[J].情报科学, 2021, 39(2): 185-192.
- [14] 中国科技资源共享网. 国家科学数据中心[EB/OL]. [2023-04-05]. <https://escience.org.cn/data-center>.
- [15] 中国科学院科学数据中心网. 科学数据中心体系[EB/OL]. [2023-04-05]. <https://www.casdc.cn>.
- [16] 中国政府网. 国务院办公厅关于印发科学数据管理办法的通知[EB/OL]. (2018-03-17)[2023-02-10]. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.
- [17] 储节旺, 夏莉. 嵌入生命周期理论的科学数据管理体系构建研究: 牛津大学为例[J].现代情报, 2020, 40(10): 34-42.
- [18] 胡海鹏, 袁永. 我国科技决策智库体系及内部运行机制研究[J].科技管理研究, 2020, 40(4): 34-39.
- [19] 王卷乐, 石蕾, 徐波, 等. 我国科学数据标准体系研究[J].中国科技资源导刊, 2020, 52(5): 45-51, 77.
- [20] 司莉, 贾欢. 科学数据的标准规范体系框架研究[J].图书馆, 2016(5): 5-9.
- [21] 王卷乐, 孙九林. 地球系统科学数据共享标准规范体系研究与应用[J].地理科学进展, 2009, 28(6): 839-847.
- [22] 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): 11-17.
- [23] 刘晓娟, 孙馒莉. 生命周期视角下科学数据安全分级管理实践与启示[J].情报理论与实践, 2023, 46(3): 68-74.
- [24] 智峰, 田锋, 赵若凡. 计量科学大数据分级分类[J].大数据, 2022, 8(1): 60-72.
- [25] 何琳, 常颖聪. 国内外科学数据出版研究进展[J].图书情报工作, 2014, 58(5): 104-110.
- [26] 冯丽妃. 生物信息“数据孤岛”是如何打破的[N].中国科学报, 2022-03-17(1).
- [27] 倪思洁. 打造空间科学数据仓储的中国品牌[J].科技传播, 2022, 14(6): 13-14.