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.DOI:10.11867/j.issn.1001-8166.2019.03.0306.[Wang Juanle, Wang Mingming, Shi Lei, 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. DOI:10. 11867/j. issn. 1001-8166. 2019. 03. 0306.]

科学数据管理态势及其对我国地球科学领域的启示*

王卷乐[1,5],王明明[1,2],石 蕾[3],高孟绪[3],陈明奇[4],
郑晓欢[4],王 超[3],王玉洁[1]
(1.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;2.山东理工大学 建筑工程学院,山东 淄博 255049;3.国家科技基础条件平台中心,北京 100862;4.中国科学院办公厅,北京100864;5.江苏省地理信息资源开发与利用协同创新中心,江苏 南京 210023)

摘 要:科学数据是重要的科技基础条件和国家战略资源。随着大数据时代的到来,全球各国纷纷将科学数据管理纳入国家发展战略。为了完善我国科学数据管理并抓住这个大数据发展机遇,国务院办公厅于2018年3月正式印发《科学数据管理办法》。在这一新的国际、国内数据管理政策环境下,分析归纳了全球11个方面的科学数据管理态势,并剖析了发达国家科学数据中心发展模式,针对我国地球科学领域科学数据管理,提出了科学数据政策制定、科学数据中心建设、科学数据生命周期管理、数据机构能力建设、科学数据认证、数据出版、科学数据回报机制、数据综合集成和安全管理等方面的启示与展望。

关 键 词:科学数据管理;地球科学;数据共享;科学数据中心;大数据
中图分类号: P3 **文献标志码**:A **文章编号**:1001-8166(2019)03-0306-10

1 引 言

科学数据是信息时代传播速度最快、影响面最宽、开发利用潜力最大的战略性、基础性科技资源[1]。随着大数据时代的到来,海量科学数据不断产生,以数据驱动为特征的科学研究方法发生了重要变革[2,3]。科学发现越来越依赖于对海量数据的集成和分析,科学研究水平不仅仅取决于科研人员的水平,也越来越多地取决于对数据的积累以及将数据转换为信息和知识的能力[4~7]。科学数据已成为科技创新、经济发展和相关决策活动不可缺失的基础科技支撑条件,被公认为继物质和能量之后的第三类资源,成为重要的科技基础条件和国家战略资源[8,9]。

全球各国纷纷将数据开放管理纳入到本国发展战略。发达国家注重科学数据的开放与重用,建立了数据和信息共享法律和政策,在国家科学数据管理框架方面形成优先领域布局和多样化的运行机制[10]。我国也紧跟国际趋势,将大数据和数据共享上升为国家战略。2015年9月,国务院印发《促进大数据发展行动纲要》[11],提出"积极推动由国家公共财政支持的公益性科研活动获取和产生的科学数据逐步开放共享"。为进一步加强和规范科学数据管理,保障科学数据安全,提高开放共享水平,更好支

收稿日期:2018-12-19;**修回日期**:2019-02-27.
* **基金项目**:国家自然科学基金项目"地球科学领域科学数据管理策略研究"(编号:41842061);中国科学院信息化十三五专项"2018年院网信发展战略专题研究"(编号:XXH13511-02)资助.
作者简介:王卷乐(1976-),男,河南洛阳人,研究员,主要从事科学数据共享、地理信息系统与遥感应用研究.**E-mail**:wangjl@igsnrr.ac.cn

撑国家科技创新、经济社会发展和国家安全,国务院办公厅于2018年3月印发《科学数据管理办法》[12]。这是我国首个国家层面出台的科学数据管理办法,为我国科学数据工作确定了行动纲领,对各个领域和学科的科学数据管理都将产生直接影响。

地球科学研究是典型的数据密集型研究,其在解决科学和应用问题的过程中需要大量的科学数据支撑,同时又在相关科研活动中不断产出新的衍生数据和产品[13,14]。这些不断随着科学研究产出的科学数据及其衍生产品是重要的科学数据资源。与发达国家相比,我国缺乏战略定位清晰、有效的地球科学领域科学数据积累和共享策略[15,16],许多立足于我国的卓越科研工作还在利用来自于欧美等国的科学数据中心的资料,产出的科学数据也多以学术论文、数据论文、可获取数据集等方式存储在国外科学数据中心,尚没有形成我国自主的、可持续的科学数据管理策略。

针对这一国际、国内新的科学数据管理政策和环境,本文拟通过国际科学数据管理态势和发达国家科学数据中心实例调研,重点分析优势科学数据中心和公共数据存储机构的建设和发展经验,结合最新的科学数据管理与共享趋势,提出对我国地球科学领域科学数据管理的启示。

2　国际科学数据管理态势

结合科技部国家科技基础条件平台中心组织的国家科学数据发展报告编制工作以及中国科学院科研信息化调研工作,先后对发达国家的科学数据管理政策和科学数据中心建设进行了系统调研,初步获得的科学数据管理态势如下。

(1)　*在立法层面确立科学数据管理的基本原则*

如何让越来越多的数据资源在全社会流动起来,最大限度地发挥数据作为资源的作用,同时规范数据在管理和社会流动中的行为,是科学数据管理的核心。美国在20世纪最后10年确立了在国家层面上建设国有科学数据和信息全社会共享环境的战略部署,原则上除危及国家安全、影响政府政务和涉及个人隐私的数据和信息以外的国有(公共领域)数据和信息全部实施"完全与开放"(full and open)的共享国策。欧盟发布《欧洲研究领域开放数据获取政策和策略》,将科学数据开放存取以政策的形式加以保证[17]。欧盟地平线2020战略发布的《开放数据:创新、增长和透明治理的引擎》要求欧盟及其成员国建立相关的法律机制并采取相应的财政措施,以推动各国在开放数据领域开展合作[18]。英国科学数据共享的指导思想和原则主要体现在《布加勒斯特宣言》、《公共资助科学数据开放获取宣言》和《网络经济的未来:首尔宣言》,并于2000年通过了《信息自由法》,从2005年1月1日起开始实施[19]。法国制定《信息科学归档文件卡片》和《自由法》[20]。日本于2001年1月生效了《信息技术基本法》[21]。

(2)　*在国家层面管理和部署科学数据工作*

欧美等发达国家已经将科学数据的持续积累和开放利用提高到了国家战略的高度进行部署,通过多年持续积累,形成了一批权威、长序列、多尺度的科学数据库,并在科研过程中发挥了重要作用。例如,英国著名的洛桑农业实验站,积累了长达160多年的土壤样品和生态试验数据,成为全世界研究人类耕作制度、施肥方式和土壤酸化演变等方面不可多得的宝贵科学财富;加拿大在全国定期开展格网化的资源调查,持续积累和提高其自然资源的管理与利用能力;Argo计划在全球范围内部署海洋浮标,用于大尺度全球气候变化观测;美国国家航空航天局(National Aeronautics and Space Administration, NASA)、美国国家海洋和大气管理局(National Oceanic and Atmospheric Administration, NOAA)和美国国立卫生研究院(National Institutes of Health, NIH)等机构支持建立的多个国家数据中心,为美国以及全球航天、大气、海洋和生命科学研究提供了重要数据资料;由发达国家主导形成的全球碳监测网络,在应对全球气候变化国际合作中发挥了关键作用等。

(3)　*围绕科学数据全生命周期加强科学数据管理*

科学数据同其他科技资源一样,具有形成、成长、成熟和衰亡的生命过程。宏观的科学数据管理贯穿整个科学数据生命周期,通常包括数据收集、数据归档、数据认证、数据加工、数据质控、数据保存、数据发布和数据共享等[22]。发达国家多针对科学数据生命周期各环节制定出台了相应管理细则予以规定和保障。例如美国国家自然科学基金会(National Science Foundation, NSF)要求在项目建议书中必须包括"数据管理"计划。英国的科学研究资助机构的科学数据管理规定涉及到数据的生命周期的各个阶段,包括:数据类型、格式、标准与数据采集方法;数据使用道德与知识产权;数据检索、共享与重用;短期保存与数据管理;数据长期保存;资金与人力支持;监督与评估。

（4） *将国家科学数据中心建设作为科学数据管理的重要手段*

美国在20世纪最后10年由NASA主导建立的地球科学领域系列国家级分布式活动存档中心（Distributed Active Archive Centers，DAACs）一直发展至今，长期为地球科学领域的国家科学数据存储、管理和发布服务。欧洲国家也在推进多个层面的科学数据中心的管理，以便为全社会提供开放的数据存储平台。Re3data.org（Registry of Research data Repositories）是由德国科学基金会（Deutsche Forschungsgemeinschaft，DFG）资助建设的全球性研究数据仓储注册系统。截至2018年9月，在该系统内注册的全球科学数据仓储/共享平台共2 088个。根据数量排名，美国1 011个，德国336个，英国294个，国际组织202个，欧盟197个，中国大陆38个，中国台湾9个，中国香港1个。截至2018年10月，国际科学理事会世界数据系统（International Science Council-Would Data System，ISC-WDS）共包含115个成员组织[23]。其中，有66个组织涉及地球科学，占比57%，包含地震学、地理信息学、土壤学、气候学、气象学、水文学、大气科学、海洋学、地球物理学、地球与环境科学、日地物理学、大气遥感学、地磁学、古气候学、地壳动力学、环境气候学、大地测量学、冰冻圈与极地学、古生态学、天体测量学、导航、地球观测、空间环境、空间大地测量和过去全球变化等。

（5） *学科领域公共数据存储是科学数据管理的有效形式*

学科领域的公共数据存储是一种具有活力的数据管理模式。这些存储以数据库、数据平台和网站的形式对外接纳数据提交者的数据，同时发布共享。例如2009年5月，美国联邦政府正式启用了Data.gov网站，该网站是美国官方的公共数据资源分享网站。Data.gov对这些数据资源的发布、采集、管理、下载和开发创建了一套科学运作体系。大学间政治社会研究联盟（Interuniversity Consortium for Political and Social Research，ICPSR）建立于1962年，是由密歇根大学主导建设的会员制联盟机构，负责收集保存社会科学和行为科学领域的数据，并为社会行为研究提供数据支持和方法培训[24]。目前共存储有超过25万份数据文件，涉及教育、老龄化、刑事司法、物质滥用和恐怖主义等21个领域，是全球最大的社会科学数据中心。地球数据观测网络（Data Observation Network for Earth，DataONE）由NSF赞助，建立于2009年，隶属于美国新墨西哥大学图书馆。DataONE是通过分布式框架和可持续网络基础设施建立的创新型环境科学数据中心，以满足人们对地球科学数据进行持续而安全的访问。Dryad数据仓储是2008年由NSF赞助成立的非盈利性会员制组织，支持存放各类数据，旨在实现对进化生物学领域期刊论文的数据的保存、发现、重用和管理的科学数据仓储。截止2018年10月，Dryad共发布24 502个和文章相关的数据包。

（6） *多手段保障科学数据生产者、服务者权益*

科学数据和信息是一种宝贵的资源，对于这种资源如何利用，涉及到很多经济学的问题。美国在这个问题上采取的基本原则是保障国家安全、政府政务和个人隐私的基础上，谁投资谁受益。严格区分投资来源并严格区分数据的产权性质是美国科学数据纳入哪一种机制运行的最主要的标准。由国家投资产生的数据（由纳税人的钱开发的数据）应该全民受益，由私营公司投资开发的数据，私营公司理所应当获得利益。欧盟数据库保护指令、英国布加勒斯特宣言和《信息自由法》等，在科学数据的产权归属、共享管理和开发利用等方面均作了明确规定[25]。欧盟出台和资助了“欧盟知识产权帮助”项目，以保护欧盟的科技创新成果及其自主知识产权。2018年5月，欧盟出台《通用数据保护条例》，以欧盟法规的形式确定了对个人数据的保护原则和监管方式[26]。

（7） *注重科学数据安全与网络安全管理*

美国注重科学数据安全与网络安全管理。从克林顿时代的网络基础设施保护，到布什时代的网络反恐，再到奥巴马时代创建的网络司令部，美国的数据安全战略经历了一个“从被动预防到主动出击”的发展过程。随着数字技术的快速发展，美国先后调整了国家信息安全政策，以巩固其在国际的领先地位，促使数据安全在国家信息安全政策中的地位不断上升。欧盟早在1996年就发布了《欧洲议会与欧盟理事会关于数据库法律保护的指令》，试图统一数据库的保护，但实际无法达到真正的统一，因此不得不分别确定数据库的版权保护和特殊权利保护的法律规范，即进行版权和特殊权利的双轨制立法，以保障科学数据共享活动的有序开展。欧盟于2007年3月22日正式通过了《关于建立欧洲信息安全社会战略的决议》[27]，标志着欧盟已经将区域的信息安全提升到社会形态的高度。2018年，经过欧盟议会长达四年的讨论，被称为史上最严数据保护法案的欧盟

《通用数据保护条例》(General Data Protection Regulation,GDPR)在5月25日生效。

(8)　科学数据中心认证

世界数据系统(World Data System,WDS)和荷兰数据认可印章(Data Seal of Approval, DSA)合作建立可信数字仓储核心认证机制(CoreTrustSeal)[28],认证机制共包括16项指南:① 仓储要把保存数据、保证数据可访问作为明确的使命;② 仓储要制定并维护用于数据访问和使用的许可,并监督数据使用者遵从相关许可;③ 仓储要制定持续性计划,确保仓储内的数据保存和使用的可持续性;④ 仓储保证在数据的创建、访问、使用过程中,遵守相关法律法规和伦理道德;⑤ 仓储要具备充足的资金和专业的工作人员,制定明确的管理制度有效地践行使命;⑥ 仓储要采取相应机制确保专家指导和反馈的持续性;⑦ 仓储必须确保数据的完整性和真实性;⑧ 仓储保存的数据和元数据必须符合相关标准,确保数据对用户的易懂性和关联性;⑨ 仓储要采用文件化的程序管理数据的归档和存储;⑩ 仓储要承担长期保存数据的责任,并制定相关的规划与文件;⑪ 仓储要具备相应的专业知识来处理技术数据和元数据,并提供足够的信息供用户进行数据质量评估;⑫ 仓储的数据归档要根据明确清晰的数据周期工作流进行;⑬ 仓储允许用户发现数据,并以合理方式持久引用;⑭ 仓储允许重用数据,并确保有标准的元数据用于对数据的理解和使用;⑮ 仓储在支持的操作系统和其他核心基础设施软件上,使用适合于其用户群体的硬件和软件技术;⑯ 仓储的技术基础设施要能够为其数据、产品、服务和用户提供保护。

(9)　科技计划项目产出的科学数据汇交

科技计划项目是产生科学数据的重要源泉,及时汇交和共享这些数据资源,既是国家科技投入的直接效益体现,也是促进这些数据被更好地归档、存储、共享和开发利用的重要途径[29]。欧美发达国家早在20世纪90年代就已经开始制定相应的数据汇交政策,陆续开展了实质性的科技计划项目数据汇交。NSF要求2011年1月18日之后,所有提交到NSF的项目申请书必须包含一份不超过2页的“数据管理计划”附件[30],并于2015年发布《NSF公共获取计划》要求把NSF资助的研究成果(包括研究数据)在首次发表后的12个月内通过存储方式保存下来,并实现开放共享与利用[31]。NOAA数据共享政策发布于2011年10月,要求项目资助产生或者衍生的环境数据和信息应该及时地(通常不迟于数据生成后的2年)可视化、可用化。我国科技部于21世纪初开始研究和制定数据汇交政策,2008年首次启动了国家重点基础研究发展计划(973计划)资源环境领域项目数据汇交[32],2011年在人口健康和农业领域开展“十一五”科技计划项目数据汇交[33],随后于2013年启动了科技基础性工作专项项目数据汇交与规范化整编工作[34,35]。国家自然科学基金委员会也开展了数据汇交实践并建立了中国西部环境与生态科学数据中心[36,37]。

(10)　利用权威数据中心推进科学数据积累、管理与利用

发达国家积极依靠权威性科学数据中心,持续整合和汇聚全球科学数据资源,并逐渐形成标准化的科学数据收集、管理和存储解决方案。例如,全球生物多样性信息网络(Global Biodiversity Information Facility, GBIF)是目前全球最大的生物多样性信息服务机构,该组织通过和种子基金合作等各种途径促进生物多样性原始数据(primary data)的共享,已形成一个面向全世界用户的、关于全球生物多样性的综合性信息服务系统。GeneBank已经成为世界权威的基因序列登记库,并被科学共同体所接受,发表学术论文往往需要提供基因登记号。

(11)　学术期刊和数据出版推进科学数据共享

受大数据影响,不仅数据中心在积极吸纳数据资源,多个国际学术期刊也正在通过各种方式整合科学数据。例如,*Nature*杂志在线发布和共享了80多种期刊论文和相关数据服务,每月全球有数以百万计的科研人员对其进行浏览和访问。其中,科学数据发布(Scientific Data)是*Nature*重要的数据服务形式。Figshare是一个在线数据知识库,用户可以存储、分享和发现科研数据。目前已与John Wiley和Sage等多个国际出版商达成合作,支持有意公开分享其数据的作者,对现有的期刊工作流程和文章出版物进行面向数据共享的整合,推出新的期刊论文出版模式。国内也先后涌现出《中国科学数据》、《全球变化数据学报》和*Big Earth Data*等数据期刊,推进科研数据共享。

3　国际地学领域科学数据中心示例

美国在国家科学数据中心建设上具有全球代表性,尤以NASA的地球科学领域数据中心管理最为典型。NASA的数据管理计划对NASA所有的科

研项目和研究人员提出数据管理计划(Data Management Plan, DMP)要求。地球观测系统数据和信息系统(Earth Observing System Data and Information System, EOSDIS)是NASA地球科学数据系统的核心。它提供端到端功能,用于管理NASA各种来源(卫星、飞机、现场测量和各种其他程序)的地球科学数据。EOSDIS被设计成分布式系统,由分布在美国各地的DAACs的主要设施构成。NASA下属的12个科学数据存档中心分工清晰,具有业务衔接关系,且不重复。以下简要列举各数据中心的依托机构和主要归档数据内容。图1为各归档中心的空间布局。

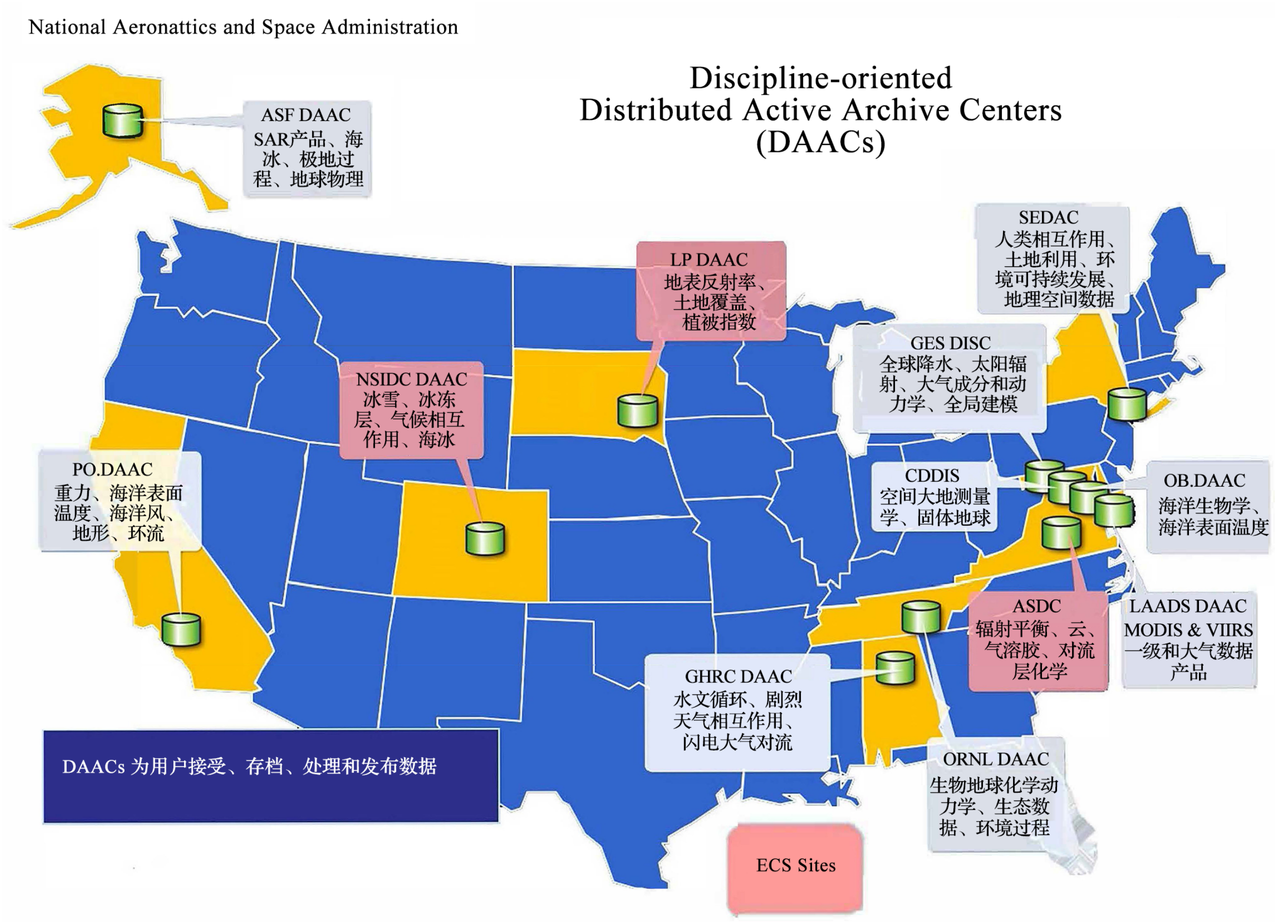

图1 美国地学领域科学数据中心的布局(https://earthdata.nasa.gov/about/daacs)

Fig.1 The location of the scientific data center in the field of geosciences in the United States (https://earthdata.nasa.gov/about/daacs)

阿拉斯加卫星设施(Alaska Satellite Facility, ASF)DAAC位于费尔班克斯的阿拉斯加大学地球物理研究所,归档极地轨道卫星和机载传感器获取、处理、归档和传播的合成孔径雷达(Synthetic Aperture Radar, SAR)数据。

大气科学数据中心(Atmospheric Science Data Center, ASDC)位于NASA兰利研究中心,负责在辐射预算、云、气溶胶和对流层化学领域的NASA地球科学数据的处理、存档和传播。

地壳动力学数据信息系统(Crustal Dynamics Data Information System, CDDIS),归档和分发主要的全球导航卫星系统(GNSS,GPS和GLONASS)、激光测距、超长基线干涉测量和多普勒轨道成像和无线电定位综合卫星数据,为地球物理学研究提供支撑。

全球水文资源中心(Global Hydrology Resource Center, GHRC)DAAC是NASA马歇尔太空飞行中心(Marshall Space Flight Center, MSFC)和阿拉巴马汉茨维尔大学(University of Alabama in Huntsville, UAH)的合作伙伴,位于UAH校园的国家空间科学技术中心(National Space Science and Technology Center, NSSTC),在灾害性天气、管理动力学、物理过程和相关应用方面提供全面的数据与知识增值

服务的活动式存档,重点关注雷电、热带气旋和风暴危害。

戈达德地球科学数据和信息服务中心(Goddard Earth Sciences Data and Information Services Center, GES DISC)是NASA的降水和水文学存档中心,存档大气组成和动态遥感数据和信息,提供现代回顾性分析研究与应用(Modern Era Retrospective-Analysis for Research and Applications, MERRA)数据同化数据集、北美土地数据同化系统(North American Land Data Assimilation System, NLDAS)和全球土地数据同化系统(Global Land Data Assimilation System, GLDAS)数据产品。

土地处理分布式活动存档中心(Land Processes DAAC, LP DAAC),作为美国地质调查局(U.S. Geological Survey, USGS)和NASA之间的合作伙伴,是NASA EOSDIS的组成部分。LP DAAC成立于1990年,用于处理NASA的陆地过程数据产品,为集成地球系统的跨学科研究提供重要贡献。LP DAAC采集、处理、归档和分发与2个地球观测系统(Earth Observing System, EOS)传感器:高级星载热发射反射辐射计(Advanced Spaceborne Thermal Emission and Reflection Radiometer, ASTER)和中分辨率成像光谱仪(Moderate-resolution Imaging Spectroradiometer, MODIS)收集的土地过程相关的数据产品,以及由特定首席调查员创建的任务衍生产品。

一级和大气档案和分配系统(Level 1 and Atmosphere Archive and Distribution System, LAADS) DAAC为全球NASA Terra, Aqua MODIS和国家极地轨道伴随卫星(Suomi national polar-orbiting partnership, Suomi NPP)可见红外成像辐射计套件(Visible Infrared Imaging Radiometer Suite, VIIRS)的科学以及应用提供服务。

国家冰雪数据中心(National Snow and Ice Data Center, NSIDC) DAAC归档和发布数字和模拟冰雪数据,保存关于积雪、雪崩、冰川、冰盖、淡水冰、海冰、地面冰、多年冻土、大气冰、古冰川学和冰芯的信息。

橡树岭国家实验室(Oak Ridge National Laboratory, ORNL)DAAC存储包括与生物地球化学和生态系统过程相关的地面和遥感测量数据,数据源于NASA赞助的野外调查、通量塔、地球观测系统卫星、相关模型输入和输出、模型源代码,以及其他对全球变化研究领域有价值的生物地球化学和生态动力学数据。

海洋生物学分布式活动存档中心(Ocean Biology DAAC, OB.DAAC),NASA的海洋生物处理组(Ocean Biology Processing Group, OBPG)作为分布式活动存档中心,负责归档卫星海洋生物学数据产品或由EOSDIS收集的数据。

物理海洋学分布式活动存档中心(Physical Oceanography DAAC, PO.DAAC)隶属美国航空航天局喷气推进实验室,归档、发布和提供NASA卫星海洋学的科学信息服务,包括天气预报、气候研究和海洋数据管理,提供关于全球海洋的物理过程和状况的数据(如海洋风、温度、地形、盐度和海洋环流等)。

社会经济数据和应用数据中心(Socioeconomic Data and Applications Data Center, SEDAC),由国际地球科学信息网络中心(Center for International Earth Science Information Network, CIESIN)负责,隶属哥伦比亚大学地球研究所,主要收集地球科学和社会经济数据和信息,目标是帮助科学家、决策者和公众更好地了解人与环境之间变化的关系。

4 地学领域科学数据管理的启示

随着我国对科学数据的日益重视,我国科学数据管理取得了显著进展。但与发达国家科学数据发展态势相比,我国在国家层面的科学数据政策体系不完备;科学数据中心建设尚显薄弱,数据资源较为分散,影响力明显不足;科学数据仍缺乏规范化的数据持续积累,在某种程度上也导致了我国科学数据的流失。我国地球科学研究近年来如火如荼地开展,更凸现出基础科学数据支撑不足的问题。结合以上调研和趋势分析,提出对我国地球科学数据管理的若干启示如下。

(1) 国家层面的科学数据管理政策体系与配套细则

欧美等发达国家制定科学数据管理的法律和制度,从行政、经济、权益等角度从根本上实现国家投资科学数据的汇集、管理与开放服务。科学数据管理的政策框架清晰,且实施方式灵活,既保障科学数据管理的合法性又不约束单一的科学数据汇集、管理和开放服务模式。尽管我国已经制定了《科学数据管理办法》,但其在行业、部门、领域的辐射力度有限,对上需要有法律层面的制度保障,对下需要有更具实施操作性的配套细则。例如,针对地球科学领域或相关部门地学数据管理的指导意见和管理细则。

（2） 地学国家科学数据中心和学科领域公共存储协同发展

各类国家级科学数据管理机构和开放的学科领域公共存储是数据管理的主体。欧美注重科学数据管理的行业领域分工协作，避免重复。地学领域国家科学数据中心按学科领域组织，分工清晰，特色鲜明。领域自建的非国家直接投资的各类公共数据存储发达，面向市场竞争，形成长效机制。二者协同发展，并可能互为影响甚至相互转换。其在发展中的共性有3点：一是研究型数据中心建设重视数据编目与永久性保存，二是数据管理机构重视数据增值开发与应用服务，三是致力于建成可持续的运行模式并遵从学科发展和市场规律。

（3） 科学数据管理的生命周期清晰，形成闭环

科学数据的生命周期特征清晰，从数据产生、存储、重用的各个阶段职责分明。普遍对国家投资的各类科技计划项目所产生的科学数据实施数据管理计划、数据汇交归档和开放服务的流程管理，严格从法律、资金以及学术道德等层面进行约束，形成闭环。科学数据的汇交、管理与共享服务融为一体。同时，基于信息技术建立跨学科、跨地域的科学数据服务系统。典型的像美国的国家环境信息中心，利用将气候、海洋、地球物理数据中心集成为虚拟的国家数据中心，开展覆盖全球的冰雪、大气、海洋、卫星遥感等领域的数据服务，形成业务上的闭环。

（4） 科学数据管理机构优胜劣汰、多机制促进自我生长能力

科学数据管理机制存在着国家法律约束、行业领域布局统筹、科学政策引导、市场驱动等多种科学数据管理机制。国家支持、会员收费、市场赢利、众筹自愿等多种模式共存。各类数据管理机构遵循成本预算和市场压力，优胜劣汰现象普遍，在这一压力下各类科学数据管理机构主动加强自我生长能力。为此，地学领域各科学数据管理机构自我发展的能动性要更为突出，积极面向用户提供数据服务，注重在用户群体中的声誉，提高自身在用户群体中的粘着性，进而提高在学术团队和社会中的生存能力。

（5） 科学数据中心认证机制初步建成

DSA的分层认证、逐步完善的认证流程对我国建立自主科学数据仓储认证机制有重要借鉴意义。可参考DSA认证的经验，建立我国自主的科学数据仓储评价机制和认证体系，不仅有利于我国科学数据仓储认证事业的发展，完善我国科学数据政策。目前经过认证的DSA社区中亚洲的地学数据中心只有一家（http://www.igsnrr.cas.cn/xwzx/zhxw/201902/t20190218_5241107.html）。借助DSA认证经验，探索适合我国可信赖数据仓储建设发展的标准，一方面便于推动和提升我国可信赖数据仓储的建设发展；另一方面也使我国有更多的机会在国际相关领域发声和交流，促进我国科学数据管理的国际化与影响力。

（6） 数据出版的驱动机制有力支持科学数据管理与共享

欧美在科学研究数据管理方面、数据出版以及科技期刊要求提供和开放数据的做法，有力促进科学数据的管理与开放。这一兼顾政策和学术道德的数据管理做法，甚至比国家科技计划项目数据汇交管理的影响力和执行力都显著。由于科学数据出版面向全球，这也意味着发达国家在科学数据管理的能力上也因此而影响全球，也是快速汇聚全球科学数据资源的利器。由于能够较好地解决数据版权问题，当前地学领域的数据出版正快速发展。期刊论文关联数据与数据出版直接发布数据2种方式将快速驱动地学科学数据共享进程。

（7） 科学数据管理的反哺和回报机制有效

欧美发达国家利用信息技术对科学数据管理和开放服务的效益进行量化和追踪，协调解决数据提供者、管理者、使用者之间的利益关注点，使得科学数据管理成效能够开放透明地反哺和回报各利益相关方。例如，ICPSR利用信息技术对科学数据管理和开放服务的效益进行量化和追踪。ICPSR当前引用率较高的数据引用次数高达850余次。这一数据引用追踪机制是对科学数据资源拥有者和提交者的极大回报，也在客观上促进优质科学数据资源的社会推广和科学界评价。除了引用率，地学领域科学数据共享的绩效评价机制仍然是个关注的重点。

（8） 加强面向问题导向的地球系统科学数据综合集成

以问题为导向的数据共享平台建设有利于打破学科界限，在高度综合地球科学整体研究对象的基础上提出数据整合集成的学术思想和方向，且易与大型国际/国家科学计划相结合[38]，促进数据的产生、集成和应用。例如，CIESIN在解决地球系统科学数据管理中强调了人类活动的影响，突出人类活动与地球环境关系的数据资源建设，这也是其能够快速着眼于可持续发展应用的一个特点。我国在地球系统科学与可持续发展方面存在的问题多而

复杂，加强问题导向的数据资源整合集成，也是当前我国地学领域科学数据资源管理的紧迫需求。

（9）　注重地学领域科学数据安全管理

总体上，我国在科学数据上取得了重大进展，但是与国外发达国家相比，我国科学数据的知识产权和物理安全管理方面仍然差距明显。针对数据产权问题，对数据业者的数据知识产权作出界定和原则性规定，做好与知识产权法律制度的衔接，合理保护数据业者的数据资源权益。针对物理安全，应加强对数据采集、传输、存储、处理、使用、销毁等生命周期全过程的安全防护，构建大数据全生命周期运行保障体系。

随着国家《科学数据管理办法》的出台，越来越多的地学领域综合和分支数据平台或数据中心将得以加强建设和快速发展。这些已有的和新增的科学数据平台之间应避免重复低水平建设，加强数据平台之间的交流合作，提高地学科学数据共享的效率。在国际层面，这些科学数据平台或中心应与当前的重大国际科学计划和国家战略相结合，例如面向2030年的联合国可持续发展目标、"一带一路"倡议的重点领域和区域需求等[39,40]。通过国内、国际交流和合作，夯实我国地学数据的自身基础，充分、合理引进我们急需的国际数据资源，为更多地学领域的全球和区域性合作提供可持续的数据支撑。

参考文献（References）：

［1］ Sun Jiulin, Lin Hai. Earth System Research and Science Data［M］. Beijing: Science Press, 2009.［孙九林，林海. 地球系统研究与科学数据［M］. 北京：科学出版社，2009.］

［2］ Atkinson M, Liew C S, Galea M, *et al*. Data-intensive architecture for scientific knowledge discovery［J］. *Distributed & Parallel Databases*, 2012, 30(5/6):307-324.

［3］ Liu J, Pacitti E, Valduriez P, *et al*. A survey of data-intensive acientific workflow management［J］. *Journal of Grid Computing*, 2015, 13(4):457-493.

［4］ Cooper M M. Data-driven education research［J］. *Science*, 2007, 317(5 842):1 171.

［5］ Pentland A. The data-driven society［J］. *Scientific American*, 2013, 309(4):78.

［6］ British Academy. Data Management and Use: Governance in the 21st Century［R］. London: The Royal Society, 2017.

［7］ Editorial. Data sharing and the future of science［J］. *Nature Communications*, 2018, 9:2 817.

［8］ Guo H. Big Earth data: A new frontier in Earth and information sciences［J］. *Big Earth Data*, 2017, 1(1/2): 4-20.

［9］ Guo Huadong. Scientific big data—A footstone of national strategy for big data［J］. *Bulletin of Chinese Academy of Sciences*, 2018, 33(8): 768-773.［郭华东. 科学大数据——国家大数据战略的基石［J］. 中国科学院院刊，2018，33(8)：768-773.］

［10］ National Science & Technology Infrastructure Center. National Science Data Resource Development Report: 2016［M］. Beijing: Scientific and Technical Documentation Press, 2016.［国家科技基础条件平台中心. 国家科学数据资源发展报告：2016［M］. 北京：科学技术文献出版社，2016.］

［11］ State Council. Action Plan for Promoting Big Data Development［EB/OL］. Beijing: State Council, 2015.［2018-11-12］. http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm.［国务院. 促进大数据发展行动纲要［EB/OL］. 北京：国务院，2015.［2018-11-12］. http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm.］

［12］ General Office of the State Council. The Measures for the administration of Scientific Data［EB/OL］. Beijing: General Office of the State Council, 2018.［2018-10-30］. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.［国务院办公厅. 科学数据管理办法［EB/OL］. 北京：国务院办公厅，2018.［2018-10-30］. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.］

［13］ Wang Juanle, Yang Yaping, Zhu Yunqiang, *et al*. Data archiving progress and datatypes analysis of National Basic Research Program of China (973 Program) in resource and environment field［J］. *Advances in Earth Science*, 2009, 24(8): 947-953.［王卷乐，杨雅萍，诸云强，等. "973"计划资源环境领域数据汇交进展与数据分析［J］. 地球科学进展，2009，24(8)：947-953.］

［14］ Boulton G. The challenges of a big data Earth［J］. *Big Earth Data*, 2018, (5): 1-7.

［15］ Wang J, Sun J, Zhu Y, *et al*. A study on the organizational architecture and standard system of the data sharing network of earth system science in china［J］. *Data Science Journal*, 2013, 12: 91-101.

［16］ National Science & Technology Infrastructure Center. National Science Data Resource Development Report: 2017［M］. Beijing: Scientific and Technical Documentation Press, 2018.［国家科技基础条件平台中心. 国家科学数据资源发展报告：2017［M］. 北京：科学技术文献出版社，2018.］

［17］ Aurore Nical, Julie Caruso, Eric Archambault. Open Data Access Policies and Strategies in the European Research Area and Beyond［R］. Washington: Science-Metrix Inc, 2013.

［18］ European Commission. Open data: An engine for innovation, growth and transparent governance［J］. *Open Data Impact*, 2011, (3): 1-13.

［19］ Wang Qiaoling, Zhong Yongheng, Jiang Hong. Research on UK scientific data sharing policies and regulations［J］. *Library Journal*, 2009, 29(3):63-66.［王巧玲，钟永恒，江洪. 英国科学数据共享政策法规研究［J］. 图书馆杂志，2009，29(3)：63-66.］

［20］ Fu Xiaofeng, Li Jun, Li Jianhui. A survey on international scientific data and its open access［J］. *China Basic Science*, 2007, 9(2):30-35.［傅小锋，李俊，黎建辉. 国际科学数据的发展与共享［J］. 中国基础科学，2007，9(2)：30-35.］

［21］ Hu Zhihui. Japan promulgates the basic law on information technology [J]. *Science and Technology Policy and Development Strategy*, 2002,(1):26-27.[胡智慧. 日本颁布《信息技术基本法》[J]. 科技政策与发展战略, 2002,(1):26-27.]

[22] Ye Yujiang . Strengthen construction of science & technology infrastructure promote sharing service of scientific & technical resources[J]. *China Science & Technology Resources Review*, 2016, 48(4):1-5.[叶玉江. 加强科技基础条件建设提升科技资源共享服务水平[J]. 中国科技资源导刊, 2016, 48(4):1-5.]

[23] World Data System. Membership[EB/OL]. Tokyo: World Data System, 2018. [2018-11-30]. http://www. icsu-wds. org / community/membership.

[24] Wang Mingming, Wang Juanle, Zhao Qiang, *et al*. Experiences and enlightenment of ICPSR scientific data center development [J]. *China Science & Technology Resources Review*, 2017, 49(6):100-107.[王明明, 王卷乐, 赵强, 等. ICPSR科学数据中心的建设经验与启示[J]. 中国科技资源导刊, 2017, 49(6):100-107.]

[25] Wang Yi, Hua Xia, Wang Jianmei. Analysis of the domestic and foreign scientific data management and sharing[J]. *Science & Technology Progress and Policy*, 2013, 30(14):126-129.[王祎, 华夏, 王建梅. 国内外科学数据管理与共享研究[J]. 科技进步与对策, 2013, 30(14):126-129.]

[26] European Union. General Data Protection Regulation[EB/OL]. United Kingdom: Official Journal of the European Union, 2016. [2018-11-30]. https://gdpr-info.eu/.

[27] Guo Chuntao. EU information network security law regulation and its reference significance[J]. *Netinfo Security*, 2009,(8):27-30.[郭春涛. 欧盟信息网络安全法律规制及其借鉴意义[J]. 信息网络安全, 2009,(8):27-30.]

[28] Han Xuehua, Wang Juanle, Shi Lei, *et al*. Identification and its inspiration for Netherlands' Data-Seal-of-Approval in scientific data repositories[J]. *China Science & Technology Resources Review*, 2018, 50(1): 14-19.[韩雪华,王卷乐,石蕾,等. 荷兰数据认可印章科学数据仓储认证及启示[J]. 中国科技资源导刊, 2018, 50(1): 14-19.]

[29] Wang Juanle, Sun Jiulin, Yang Yaping, *et al*. Data archiving practice and consideration of national basic research program of China (973 Program) in resource and environment field[J]. *China Science & Technology Resources Review*, 2011,43(3):1-5.[王卷乐,孙九林,杨雅萍,等. 973计划资源环境领域项目数据汇交实践与思考[J]. 中国科技资源导刊,2011,43(3):1-5.]

[30] Si Li, Xing Wenming. Scientific data management and sharing policies in foreign countries: Investigation and inspiration to us [J]. *Information and Documentation Services*, 2013, 34(1):61-66.[司莉, 邢文明. 国外科学数据管理与共享政策调查及对我国的启示[J]. 情报资料工作, 2013, 34(1):61-66.]

[31] Zhang Xiaoqing, Sheng Xiaoping. A review of foreign policies on open data and data sharing of scientific data[J]. *Library Tribune*, 2018,38(8):147-154.[张晓青,盛小平. 国外科学数据开放共享政策述评[J]. 图书馆论坛,2018,38(8):147-154.]

[32] Lin Hai, Wang Juanle. The data archiving work in resource and environment field of national basic research program of China (973 Program) has been officially launched[J]. *Advances in Earth Science*, 2008, 23(8): 895-896.[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.]

[33] Li Na, Gao Baihong. Practice and relfect on delivery of the S & T project resource in the personal and healthy[J]. *China Science & Technology Resources Review*, 2015, 47(5): 63-67.[李娜, 高百红. 人口健康科技项目资源汇交实践与思考[J]. 中国科技资源导刊, 2015, 47(5): 63-67.]

[34] Zhu Yunqiang, Sun Kai, Yang Yaping, *et al*. Data resources collection and reorganization for national special program on basic works for science and technology of China[J]. *China Science & Technology Resources Review*, 2017,49(5):12-20.[诸云强,孙凯,杨雅萍,等. 科技基础性工作数据资料的汇交与整编[J]. 中国科技资源导刊,2017,49(5):12-20.]

[35] Yang Jie, Song Jia, Zhu Yunqiang, *et al*. Construction of special data archiving and sharing platform for the science and technology basic work [J]. *China Science & Technology Resources Review*, 2017,49(5):52-59,67.[杨杰,宋佳,诸云强,等. 科技基础性工作专项数据汇交共享平台建设[J]. 中国科技资源导刊,2017,49(5):52-59,67.]

[36] Li Xin, Nan Zhuotong, Wu Lizong, *et al*. Environmental and ecological science data center for West China: Integration and sharing of environmental and ecological data[J]. *Advances in Earth Science*, 2008, 23(6):628-637.[李新,南卓铜,吴立宗,等. 中国西部环境与生态科学数据中心:面向西部环境与生态科学的数据集成与共享[J]. 地球科学进展,2008, 23(6):628-637.]

[37] Nan Zhuotong, Li Xin, Wang Liangxu, *et al*. Design and implementation of the online data sharing portal of environmental and ecological science data center for West China[J]. *Journal of Glaciology and Geocryology*, 2010,32(5):970-975.[南卓铜,李新,王亮绪,等. 中国西部环境与生态科学数据中心在线共享平台的设计与实现[J]. 冰川冻土,2010,32(5):970-975.]

[38] Zhao Zuoquan. The frontier of Earth science: From disciplinary guidance to problem guidance—Characteristics, comparison and reflection of the frontiers of Earth science in the United States and China[J]. *Science & Technology Review*, 1994,8:13-15.[赵作权. 地球科学前沿走向:从学科导向到问题导向——美、中两国地球科学前沿的特点、比较与思考[J]. 科技导报,1994,8:13-15.]

[39] Wei Yanqiang, Li Xin, Gao feng, *et al*. The United Nations Sustainable Development Goals (SDG) and the response strategies of China[J]. *Advances in Earth Science*, 2018, 33(10):1 084-1 093.[魏彦强, 李新, 高峰, 等. 联合国2030年可持续发展目标框架及中国应对策略[J]. 地球科学进展, 2018, 33(10):1 084-1 093.]

[40] Deng Wei, Zhao Wei, Liu Bintao, *et al*. Water security and the countermeasures in South Asia based on the "Belt and Road" initiative[J]. *Advances in Earth Science*, 2018,33(7):687-701.[邓伟, 赵伟, 刘斌涛, 等. 基于"一带一路"的南亚水安全与对策[J]. 地球科学进展, 2018, 33(7):687-701.]

The Situation of Scientific Data Management and Its Enlightenment to Earth Sciences of China*

Wang Juanle[1,5], Wang Mingming[1,2], Shi Lei[3], Gao Mengxu[3], Chen Mingqi[4], Zheng Xiaohuan[4], Wang Chao[3], Wang Yujie[1]

(1.*State Key Laboratory of Resources and Environment Information System*, *Institute of Geographic Sciences and Natural Resources Research*, *Chinese Academy of Sciences*, *Beijing* 100101, *China*; 2.*School of Civil and Architectural Engineering*, *Shandong University of Technology*, *Shandong Zibo* 255049, *China*; 3.*National Science and Technology Infrastructure Center*, *Beijing* 100862, *China*; 4.*General Office*, *Chinese Academy of Sciences*, *Beijing* 100864, *China*; 5.*Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application*, *Nanjing* 210023, *China*)

Abstract: Scientific data is an important scientific and technological basis and national strategic resource. With the coming of the era of big data, countries all over the world have added scientific data management into their national development strategies. In order to improve the scientific data management in our country and make full use of the development opportunity from big data, the General Office of the State Council officially issued "The Measures for the Administration of Scientific Data" in March, 2018. In the new international and domestic policy on data management, this study analyzed and summarized 11 aspects of international situation of scientific data management, and dissected the development models of developed countries' scientific data centers deeply. And also, the study put forward new ideas for the policy-making of scientific data, the construction of scientific data centers, the life cycle management of scientific data, the capacity building of data organizations, the authentication of scientific data, the publication of data, the return mechanism of the data and the comprehensive integration, and the safety management of the data in the scientific data management of Earth sciences in China.

Key words: Scientific data management; Earth science; Data sharing; Scientific data center; Big data.

* **Foundation item:** Project supported by the National Natural Science Foundation of China "Research on scientific data management strategy in Earth Science" (No. 41842061); The 13th Five-Year Informatization Plan of Chinese Academy of Sciences "Research on the development strategy of the network information of CAS in 2018" (No. XXH13511-02).

* **Corresponding author:** Wang Juanle(1976-), male, Luoyang City, He'nan Province, Professor. Research areas include scientific data sharing, geographic information systems and remote sensing applications. **E-mail**: wangjl@igsnrr.ac.cn