

# 2020年地球数据科学与共享热点回眸

王卷乐<sup>1,2</sup>, 王玉洁<sup>1</sup>, 张敏<sup>1,2</sup>, 蒋涵<sup>1,3</sup>, 洪梦梦<sup>1,4</sup>, 李姝晗<sup>1,5</sup>

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. 中国科学院大学, 北京 100049

3. 江苏海洋大学海洋技术与测绘学院, 连云港 222005

4. 山东理工大学建筑工程学院, 淄博 255049

5. 防灾科技学院, 三河 065201

**摘要** 地球数据科学是地球科学与数据科学的结合, 是地球科学研究与数据驱动科学发现范式转变的前沿交叉领域。在地球大数据研究快速兴起的背景下, 从地球数据大科学计划、科学数据治理和地球科学数据共享等方面回顾了2020年的进展。介绍了深时数字地球大科学计划、地球大数据科学工程, 以及全球数据汇聚、数据认证、数据仓储、数据政策、数据标准、数据标识等方面的进展和地球科学数据中心、数据库、数据模型、数据集成和数据出版等实践。展望未来, 充分重视科学数据治理的规则, 推动地球科学大数据计划发展, 加强建立地球科学数据生态, 促进地球科学数据领域的各方面合作, 仍然是当前的紧迫需求。

**关键词** 地球科学; 数据科学; 数据共享; 数据治理; 大数据

地球科学是认识行星地球形成和演化的一门自然科学。地球科学是以包括大气圈、生物圈(含土壤圈)、岩石圈和水圈(含冰冻圈), 乃至人类圈的各圈层及其组成部分为对象, 研究地球系统整体及各圈层之间的相互作用过程、变化、规律、机理和相互关系等, 以达到提高对行星地球的认知水平, 创

建表征行星地球的特征和动力学的理论的目的; 把获取的知识体系运用到解决人类生存与可持续发展中存在的资源和能源供给、自然灾害防治、生态环境保护等重大问题中去, 并为其提供科学依据、有效途径与技术支撑。地球科学的发展整体表现为微观更微、宏观更宏、交叉集成的姿态, 特别是进

收稿日期: 2021-01-03; 修回日期: 2020-01-08

基金项目: 中国科学院战略性先导科技专项(XDA19040501); 国家科技基础条件平台专项(2020WT22)

作者简介: 王卷乐, 研究员, 研究方向为资源环境科学数据集成与共享, 电子信箱: wangjl@igsrr.ac.cn

引用格式: 王卷乐, 王玉洁, 张敏, 等. 2020年地球数据科学与共享热点回眸[J]. 科技导报, 2021, 39(1): 105-114; doi: 10.3981/j.issn.1000-7857.2021.01.008

入 21 世纪以来,更加强调广泛采用最新观测、监测、探测、实验、分析、模拟和建立理论模型等先进技术手段,认识和理解人类赖以生存的地球和行星空间。

随着全球变化、人地关系与环境效应、地球圈层关系与机理等重大科学计划的进展,沉淀和积累了多年的科学数据资源将会起到重要的支撑作用。地球数据科学是数据科学在地球科学领域的交叉。图灵奖得主吉姆·格雷(Jim Gray)提出的第四范式,即在以实验观测为主的经验范式、以理论推演为主的理论范式、以模拟计算为主的模拟范式之后,出现的以数据为基础,联合理论、实验和模拟一体的数据密集型范式。这是一种类似方法论视角来定义的数据科学的内涵,即数据驱动科学发现<sup>[1]</sup>。

科学大数据是推动科研创新的强力引擎,是数据密集型科研时代通向未来科学研究范式的重要桥梁。地球科学大数据,作为科学大数据的重要组成部分,既具有大数据的一般性质,又具有很强的时空关联和物理关联。地球科学大数据资源建设,致力于推动地球科学的学科建设发展,同时在环境、资源、灾害等领域具有广泛的社会意义与价值。

## 1 地球数据科学计划进展

### 1.1 深时数字地球科学计划

深时数字地球科学计划(DDE)是世界上最大的地球科学联盟——国际地质科学联合会(IUGS)于2019年启动的一个科学计划。DDE旨在实现和加强地球科学中的数据驱动发现,并创建可与其他数据库互操作的深时地球科学数据中心。其愿景和使命是:“整合地球演化全球数据、共享全球化地学知识;推动地球科学的变革性发展”。DDE将建立一个遵从FAIR原则的(可查找、可访问、可交互操作和可重复使用)的数据基础设施,将现有的深时地球科学数据库连接起来,并为广大学者、研究人员、政策专业人员和公众提供量身定制的产品和服务。

DDE大数据体系的重要建设内容之一是解决地球科学中的“长尾”数据,即非结构化以及固有的在机构、大学和地球科学家个人计算机中的异构地球科学数据(图1<sup>[2]</sup>)。通过连接和协调长尾深时数据“岛屿”来改变地球科学数据共享和驱动科学发现方式,以支持与整个地球系统相关的广泛科学研究。

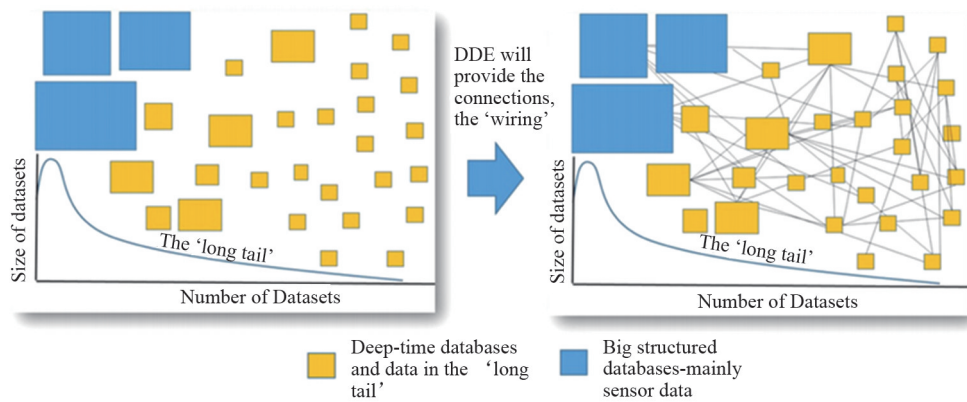


图1 DDE将提供连接深时数据源的路径

### 1.2 地球大数据科学工程

中国科学院A类战略性先导科技专项“地球大数据科学工程”(CASEarth)(以下简称“专项”)于2018年1月1日正式立项<sup>[3]</sup>。该专项致力于构建包含资源、环境、生物、生态等多个领域的大数据与云

服务共享平台,重点开展包括全景美丽中国、数字“一带一路”、三维信息海洋和时空三极环境、生物多样性与生态安全等方面的基础性相关应用研究(图2<sup>[4]</sup>),打造国际一流的数字地球科学平台,实现地球系统科学的重大突破和科学发现,推动地球大

数据领域的技术创新,实现全方位多层次的宏观决策支持和社会公众知识传播服务,为“一带一路”、数字中国、人类命运共同体和联合国可持续发展目标等提供强有力的科技支撑与决策支持。

CASEarth的数据资源规划主要包括对地观测数据、地理资源数据、生态系统地面观测数据、生物

多样性与生态安全数据、海洋数据、三极区域集成数据、“一带一路”区域集成数据、联合国可持续发展目标实现支撑数据8个主要维度<sup>[5]</sup>。依托该大数据工程,在第75届联合国大会期间,中国发布《地球大数据支撑可持续发展目标报告(2020)》,为各国加强2030年议程落实监测评估提供借鉴<sup>[6]</sup>。

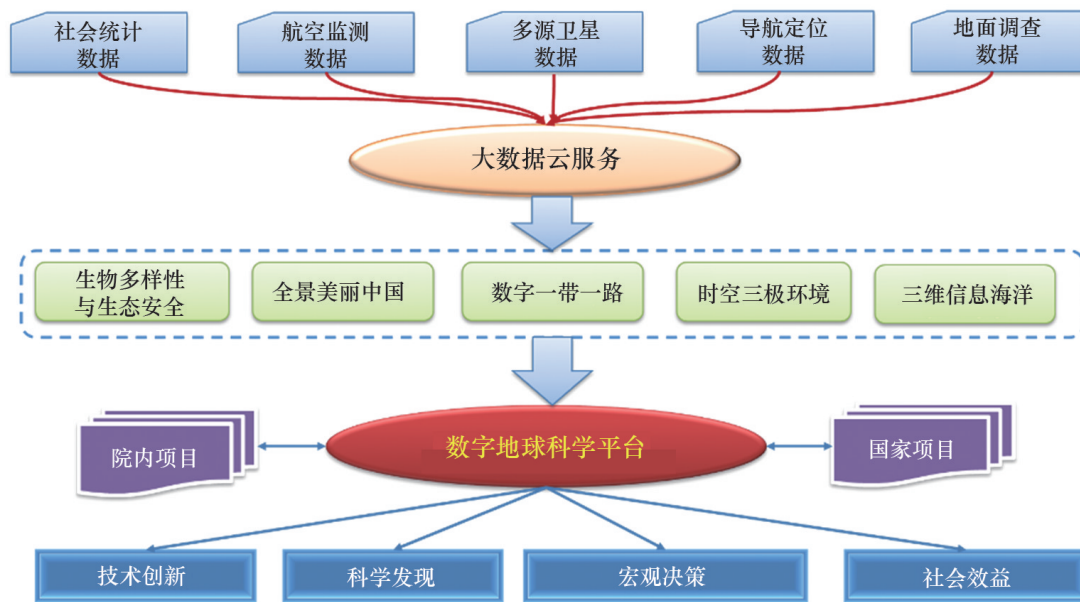


图2 地球大数据科学工程框架

## 2 科学数据治理进展

### 2.1 Data Together 行动

2020年3月, CODATA (Committee on Data of the International Science Council)、研究数据联盟 (Research data alliance, RDA)、国际科学理事会世界数据系统 (International science council-world data system, ISC-WDS) 和 GO FAIR (Global open fair) 4大国际数据组织提出 Data Together 计划(图3)<sup>[7]</sup>, 共同致力于优化全球研究数据生态系统, 并确定将在新冠肺炎疫情重大社会危机的背景下, 推动联邦基础设施以服务于数据驱动科学的新需求。Data Together 计划从以下几方面开展工作, 主要包括: 广义协议; 政策建议; 协议、技术和基础设施; 规范和标准; 社区参与和建立信任机制; 形成社区共识;



图3 国际Data Together行动

协调应对具体问题。

## 2.2 科学数据国际认证

CoreTrustSeal 认证机制是 WDS 和 DSA 合作推出的可信数字仓储核心认证机制,为有意向认证的科学数据中心提供基于 DSA-WDS 核心可信数据存储库需求目录和过程的核心级认证<sup>[8]</sup>(图4)。科学数据中心认证对科学、规范的管理复杂的科学数据中心具有重要意义。目前,全球已有 104 个国际科学数据中心通过了 CoreTrustSeal 认证。中国大陆有 5 个数据中心获得认证,分别为中国天文数据中心、可再生资源与环境世界数据中心、地质科学数据出版中心、中国地球物理学科中心和国家空间科学数据中心。其中,可再生资源与环境世界数据中心是地学领域首个通过认证的数据中心<sup>[9]</sup>。

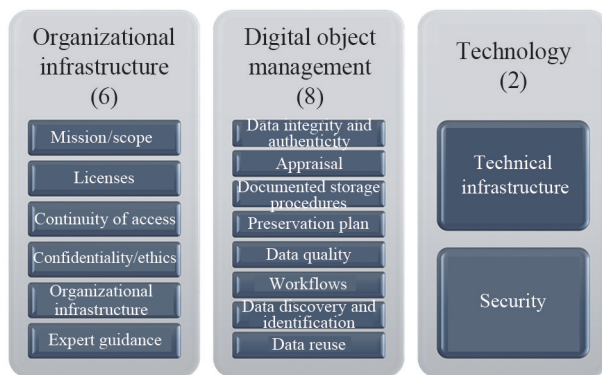


图4 CTS认证条款结构

## 2.3 科学数据仓储原则

TRUST是一个促进数据持续管理的仓储指导原则<sup>[10]</sup>,包括5个方面:1) 透明性(transparency),对于可被公开访问的证据验证的特定存储库服务和数据持有保持透明;2) 责任(responsibility),负责确保所持有数据的真实性和完整性,以及其服务的可靠性和持久性;3) 用户关注(user focus),确保数据管理规范和目标用户社区的期望得到满足;4) 可持续(sustainability),长期维持服务和保存数据;5) 技术(technology),提供基础设施和功能来支持安全、持久和可靠的服务。

TRUST原则为数据存储库参与各方开发和维护数据基础设施提供一个参考,以促进对数据的持

续管理,并使未来能够重复使用它们所持有的数据。然而,TRUST原则本身并不是目的,而是促进利益相关者间沟通的一种手段,为存储库提供展示透明度、责任、用户焦点、可持续性和技术的指导。

## 2.4 科学数据管理政策

作为数字经济时代的枢纽,建设大数据中心已经成为大势所趋。2018年3月,国务院办公厅出台了首个国家层面的《科学数据管理办法》<sup>[11]</sup>,对各个领域和学科的科学数据管理都产生直接影响。科学技术部组织建立首批 20 个国家科学数据中心(图5)。2020年6月,交通运输部办公厅发布《交通运输科学数据管理办法(征求意见稿)》<sup>[12]</sup>。2020年10月10日,中国气象局印发《气象数据管理办法(试行)》<sup>[13]</sup>,对气象部门组织开展的气象数据收集汇交、加工处理、保存使用、共享服务、安全监管等工作进行规定。

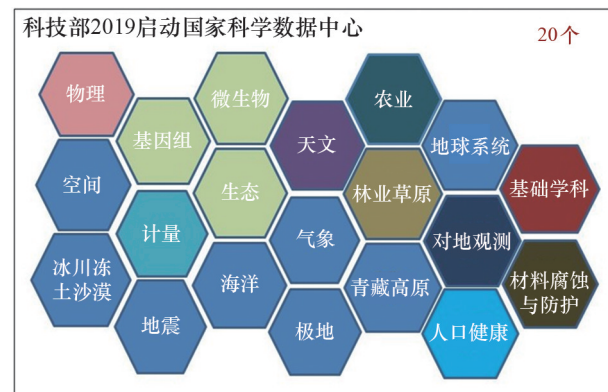


图5 中国国家科学数据中心

## 2.5 科学数据标准体系

科学数据标准体系是有序、持续推进科学数据平台和国家科学数据中心建设的必要保障。科学数据标准体系研究有利于规范科学数据管理全周期中的标准化问题,同时促进解决科技平台不同学科、领域科学数据的标准交叉、不一等问题,促进科学数据平台和数据中心的持续发展。该体系由定义与指南类基础标准,描述、采集处理、汇交、保存与维护、共享服务等通用标准及服务于行业、领域数据管理的专用标准组成<sup>[14]</sup>,如图6所示。科学数

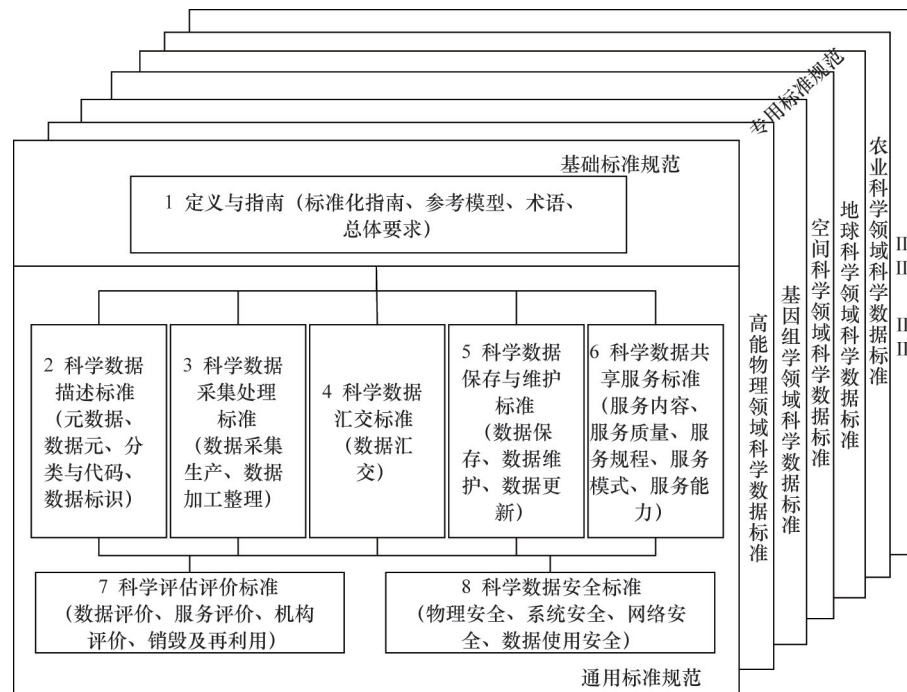


图6 科学数据标准体系参考模型

据标准体系的建立将为各类科学数据标准建议的提出和制定提供顶层框架,促进中国科学数据标准化的整体发展。

### 2.6 科学数据标识

数据标识是支撑科技资源管理及应用的重要手段,甚至成为掌控科技资源生态系统建设和发展的重要基础设施。如数字化对象识别符(Digital Object Identifier, DOI)、对象标识符(Object Identifier, OID)、中国科技资源标识(China Science and Technology Resource, CSTR)等都是科学数据通用领域常用的标识符。其中,CSTR是由国家科技基础条件平台中心为主研制的国家标准(标准号 GB/

T 32843—2016)所定义的。该标识符由中国科技资源代号缩写(CSTR)、科技资源标识注册机构代码、科技资源类型代码和内部标识符4部分组成(图7)。中国科技资源代号(4位字母码)与科技资源标识注册机构代码(5位半角字符)之间用半角符号“;”进行分割,其他各部分之间用半角符号“.”隔开。2020年12月,国家科学技术部科技基础条件平台中心正式推出了中国自主的科技资源持久化标识体系<sup>[15]</sup>,依据国家标准《GB/T 32843—2016 科技资源标识》牵头组织CSTR标识体系建设,为中国科技资源,特别是科学数据资源的安全可控、有序管理、有据使用提供了基础支撑。

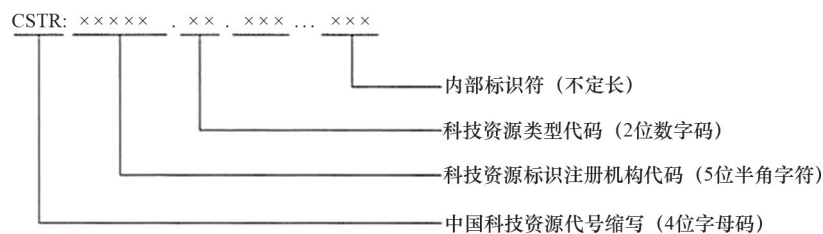


图7 CSTR 编码结构

### 3 地球数据科学数据共享进展

#### 3.1 地球科学数据中心

国际科学理事会世界数据系统(International Science Council-World Data System, ISC-WDS)为世界最大的科学数据组织。截至2020年12月,ISC-WDS共包含128个成员组织。其中,正式数据中心成员86个、网络成员11个、合作成员11个和联合成员20个。涉及多个学科领域,例如地震学、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。

Re3data.org(Registry of REsearch data REpositories)是由德国科学基金会资助建设的全球性研究数据仓储注册系统。截至2020年12月,在该系统内注册的全球科学数据仓储/共享平台共3568个。根据数量排名,美国1098个、德国429个、英国292个、欧盟278个、加拿大258个、国际组织248个、法国110个、中国大陆47个、中国台湾10个、中国香港3个。

作为地球观测领域规模最大、最具权威和影响力的政府间国际组织,地球观测组织(Group on Earth Observation, GEO)发起和推动数据枢纽建设,预期通过一个建立在云端的平台,将开放获取的数据、论文、算法、模型和计算能力融合在一起。

#### 3.2 地球科学数据库

地球科学领域数据库总体朝着多元化、标准化、智能化和一体化方向发展。在古生物学领域,过去30~40年里,随着计算机、数据库和互联网技术的快速发展,国内外涌现出大量的古生物学数据库,彼此间的目标、体系架构、数据组织方式和服务对象通常存在显著差异,呈现百花齐放的特点<sup>[16]</sup>。国内相关学者在此相关基础上结合相关实例,提出了一站式全生态链数据平台的建设设想。

在古地理学领域,面对海量的古地理数据的不

断积累,各类学会和组织发起的大数据平台建设快速发展,如CEED、TSCreator、EarthByte、Macrostrat等,这使得标准化、智能化的数字古地理的重建成为可能。数字古地理重建核心思路是:1)建立标准化的古地理学知识体系;2)利用新兴网络技术例如机器阅读来扩大数据源范围,以建立开放互动、动态更新的古地理数据库;3)建立一体化、标准化的古地理学数据质量控制体系;4)利用机器学习技术建立各类型古地理重建模型,深度挖掘数据;5)让数据以可实时更新的智能数字地图集或多维动画形式展出<sup>[17]</sup>。

在沉积学领域,国际上涌现出一大批优秀的沉积学相关的数据库,如EarthChem、FAKTS、SMAKS、DMAKS,以及各种优秀的数据集,如世界气候敏感性沉积物数据集、陆相冲积相泥质岩数据集等。这些数据库(集)尝试利用大数据的思维,从全球视野理解深时沉积物质的演化和循环过程<sup>[18]</sup>。

#### 3.3 地球科学数据模型

为对地球科学数据库中不同来源的、超大规模的地球科学数据进行高效管理,需要根据不同的需求对数据建立有效的数据模型。针对不同领域,对数据模型有不同的分类方式。就数据模型在时间顺序上的发展阶段而言,数据模型有如下4个阶段的发展历程(图8)<sup>[19]</sup>。

结构化模型,即传统数据模型,按抽象应用层次通常可分为概念数据模型、逻辑数据模型、物理数据模型。其中概念数据模型可进一步分为E-R模型和面向对象模型,逻辑数据模型可以进一步分为层次模型、网状模型、关系模型等。半结构化模型主要有XML(extensible markup language,可扩展标记语言)模型、RDF(resource description framework,资源描述框架)模型、超模型等。随着电子商务、商业智能等应用的不断发展,用户对数据库查询分析的要求越来越高,数据分析模型成为新的研究热点。数据分析模型主要包括ROLAP(relationship OLAP)模型、MOLAP(multidimensional OLAP)模型。以NoSQL(not only structured query language)

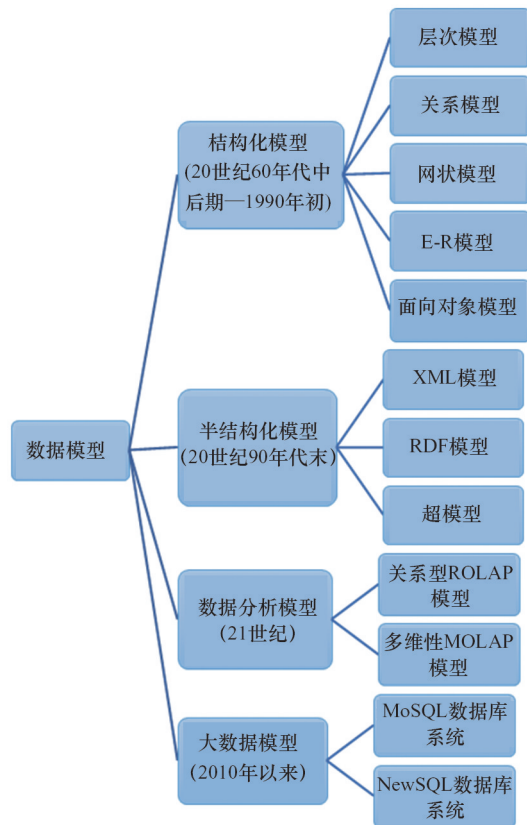


图8 按数据模型的发展历程分类

和NewSQL数据库系统为代表的大数据模型已成为当下数据模型领域主要的研究内容。NoSQL模型与传统的关系型数据库相比,具有可扩展性好、存储模式灵活、数据模型与应用层数据结构接近等优势,并越来越多地应用于大数据存储环境中, NoSQL模型主要包括Key-Value存储模型、Key-Column存储模型、Key-Document存储模型、图存储模型等。综合分析几种典型的NoSQL存储,其主要特点如表1<sup>[20]</sup>所示。

### 3.4 地球科学数据集成

地球大数据的挖掘分析涉及复杂的数据处理、模式提取和知识获取过程,需要易于使用的集成软件工具来处理大规模数据。中国科学院软件研究所正在开发一个新的地球大数据挖掘平台——Earth Data Miner。如图9所示<sup>[21]</sup>, Earth Data Miner利用云计算技术为大数据分析处理的过程提供了分布式的大数据处理引擎和机器学习引擎。科学家们可以在数据资源库查找、导入并共享公共和个人数据资源,再利用大数据处理引擎进行相关分析;另一方面,算法库提供了常用的算法,并支持科

表1 典型的NoSQL数据模型及其特征

数据模型	典型产品	主要特性
Key-Value 模型	Berkeley DB, Level DB, Memecached, Project Voldemort, Redis, Riak	聚合模型,聚合结构可以任意组织,且对用户透明
Key-Column 模型	HBase, Cassandra, Amazon Simple DB, Hypertable	双层聚合模型,有基于行和基于列两种组织方式,多个列聚合成一个列家族,并且可以自由添加列
Key-Document 模型	Couch DB, Mongo DB, Raven DB, Terastore	聚合模型,聚合结构要按照设定要求组织,用户可以看到其内部结构
图模型	Flock DB, Neo4J, Hyper Graph DB, Infnit Grapy	非聚合模型,记录小,但记录间的连接关系复杂

学家发布和共享新的算法和模型。智能分析环境提供了多种智能分析工具,如python代码IDE(integrated development environment)、Jupyter笔记本和工作流工具。同时,Earth Data Miner还提供了图表和地理信息系统的数据可视化工具,用于在线卫星图像的可视化交互分析<sup>[21]</sup>。沈志宏<sup>[22]</sup>基于现有的结构化、半结构化数据存储方式存在的不足,并结合图数据模型,发布了一种智能集成数据管理系统(Panda DB)。

该数据系统着重推进大规模数据的统一存储和查询方面的实践,并与人工智能计算方法相结合,实现透明地访问非结构化数据的内部信息的目标。

### 3.5 地球科学数据出版

随着数据密集型科研范式<sup>[23]</sup>兴起,以实体数据为核心的数据出版、数据仓储受到了许多科研机构与学者的关注,科学数据论文被提升到与学术论文同等重要的出版地位,逐步成为重要的科研成果。

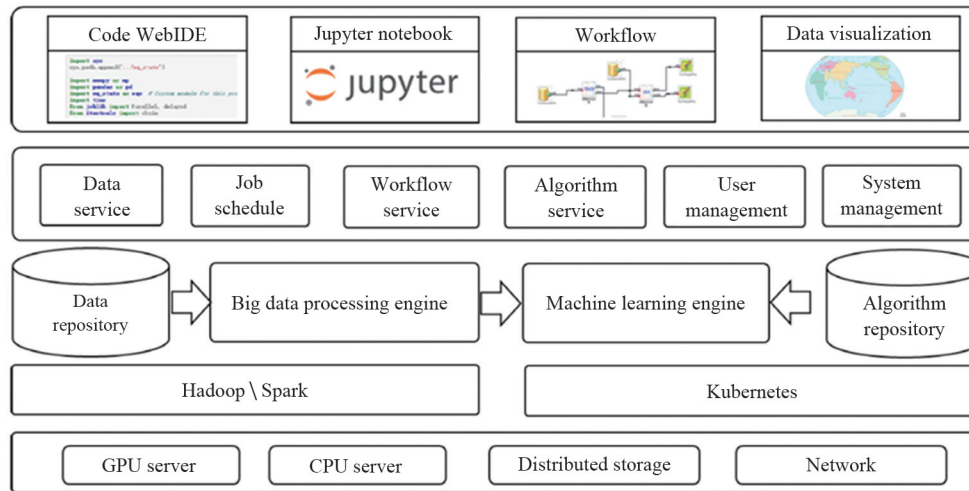


图9 Earth Data Miner 系统架构

《地球系统科学数据》(Earth System Science Data, ESSD)于2008年起发行出版,2019年其影响因子达9.197,为数据期刊类中的翘楚(图10)。目前ESSD已与德国海洋数据中心(PANGAEA)等多个数据中心合作完成数据存储<sup>[24]</sup>。2019年,美国地球物理学会(American Geophysical Union, AGU)启动期刊数据存储计划,要求旗下学术期刊论文数据必须存储于AGU认定的226家数据仓储中心<sup>[25]</sup>。中



图10 地球系统科学数据期刊(图片来源:ESSD)

国大陆境内有包括全球变化科学研究数据出版系统等15个数据仓储中心获得AGU认定<sup>[26]</sup>。《中国科学数据》《全球变化数据学报》《Big Earth Data》等数据期刊组建了较完善的数据论文审核、存储、同行评议等政策,快速推动国内数据出版。

2020年7月,国家青藏高原科学数据中心通过《Nature》数据期刊《Scientific Data》认证,是国内首个通过认证的数据仓储中心,成为《Nature》及其子刊文章投稿时可靠和便捷的数据仓储中心<sup>[27]</sup>。2020年9月,科学数据存储库(Science Data Bank, ScienceDB)被Scientific Data和Springer Nature收录到其推荐的通用型数据存储库名单,成为继Dryad、Figshare、Harvard Dataverse、OSF、Zenodo、Mendeley Data之后的第7家被收录的通用型存储库。同时,ScienceDB也是该名单中唯一一家中国自主建设维护的存储库平台<sup>[28]</sup>。

## 4 结论

地球科学数据是推动地球科学探索发现的根本驱动力。回顾了2020年地球数据科学计划进展,梳理了科学数据治理中的国际行动、数据认证、数据仓储、数据政策、数据标准体系、数据标识等方面的进展,以及在地球科学数据中心、数据库、数据模型、数据集成和数据出版方面的研究与实践。

在第四范式的指引下,地球科学数据的共享对于促进地球科学数据密集型研究具有更为现实的意义。然而,在如何让全球的地球科学数据实现融合和共享并形成高效的知识发现驱动环境方面,仍然面临着一系列挑战。充分重视科学数据治理的规则,推动地球科学大数据计划发展,加强建立地球科学数据生态,促进地球科学数据领域的各方面合作,仍然是当前的紧迫需求。

### 参考文献(References)

- [1] 程学旗, 梅宏, 赵伟, 等. 数据科学与计算智能: 内涵、范式与机遇[J]. 中国科学院院刊, 2020, 35(12): 1470-1481.
- [2] Stephenson M H, Cheng Q, Wang C, et al. Progress towards the establishment of the IUGS Deep-time Digital Earth (DDE) programme[J]. *Episodes Journal of International Geoscience*, 2020, 43(4): 1057-1062.
- [3] 中国科学院战略性先导科技专项(A类). 地球大数据科学工程[J]. 中国科学院院刊, 2018, 33(8): 760.
- [4] Guo H, Nativi S, Liang D, et al. Big earth data science: an information framework for a sustainable planet[J]. *International Journal of Digital Earth*, 2020, doi: 10.1080/17538947.2020.1743785.
- [5] 郭华东. 地球大数据: 科学工程数据共享蓝皮书(2019)[M]. 北京: 科学出版社, 2020.
- [6] Guo H. Big earth data facilitates sustainable development goals[J]. *Big Earth Data*, 2020, 4(1): 1-2.
- [7] International Science Council-World Data System. Data together final version march 2020 Final[EB/OL]. (2020-03-30) [2021-01-07]. <https://www.worlddatasystem.org/news/files/DataTogetherFinalVersionMarch2020FINAL.pdf>.
- [8] Mokrane M, Hugo W, Harrison S. WDS/DSA Certification-International collaboration for a trustworthy research data infrastructure[C]//EGU: General Assembly, 2016.
- [9] 王卷乐, 王祎, 卜坤, 等. 世界数据系统 CoreTrustSeal 数据中心认证实践——以 WDC 可再生资源与环境数据中心为例[J]. *农业大数据学报*, 2019, 1(3): 71-81.
- [10] Lin D, Crabtree J, Dillo I, et al. The TRUST Principles for digital repositories[J]. *Scientific Data*, 2020, 7: 144.
- [11] 中华人民共和国国务院办公厅. 科学数据管理办法[EB/OL]. (2018-04-02)[2021-01-02]. [http://www.gov.cn/zhengce/content/2018-04/02/content\\_5279272.htm](http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm).
- [12] 中华人民共和国交通运输部办公厅. 交通运输科学数据管理办法(征求意见稿)[EB/OL]. (2020-06-18)[2021-01-02]. [http://xxgk.mot.gov.cn/jigou/kjs/202006/t20200623\\_3398111.html](http://xxgk.mot.gov.cn/jigou/kjs/202006/t20200623_3398111.html).
- [13] 中国气象局. 风云气象卫星数据管理办法(试行)[EB/OL]. (2018-07-06) [2021-01-02]. [http://www.gov.cn/gongbao/content/2018/content\\_5350059.htm](http://www.gov.cn/gongbao/content/2018/content_5350059.htm).
- [14] 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.
- [15] 中华人民共和国科学技术部. 科技资源标识体系培训工作会在京召开[EB/OL]. (2020-12-02)[2021-01-02]. [http://www.most.gov.cn/kjbgz/202012/t20201202\\_159970.htm](http://www.most.gov.cn/kjbgz/202012/t20201202_159970.htm).
- [16] 邓怡颖, 樊隽轩, 王玥, 等. 古生物学数据库现状与数据驱动下的科学研究[J]. *高校地质学报*, 2020, 26(4): 361-383.
- [17] 张蕾, 钟瀚霆, 陈安清, 等. 大数据驱动下的数字古地理重建: 现状与展望[J]. *高校地质学报*, 2020(1): 73-85.
- [18] 蒋璟鑫, 李超, 胡修棉. 沉积学数据库建设与沉积大数据科学研究进展: 以 macrostrat 数据库为例[J]. *高校地质学报*, 2020, 103(01): 29-45.
- [19] 信俊昌, 王国仁, 李国徽, 等. 数据模型及其发展历程[J]. *软件学报*, 2019, 30(1): 142-163.
- [20] 李鸿飞, 杜溢墨, 曾熠, 等. 异构混合存储的软硬件协同数据放置策略[J]. *国防科技大学学报*, 2020, 42(2): 64-71.
- [21] Liu J, Wang W, Zhong H. Earth Data Miner: A cloud-based big earth data intelligence analysis platform[J]. *IOP Conference Series: Earth and Environmental Science*, 2020, 509(1): 012032.
- [22] 沈志宏, 赵子豪, 王海波. 以图为中心的新型大数据技术栈研究[J]. *数据分析与知识发现*, 2020, 4(7): 50-65.
- [23] 王瑞丹, 高孟绪, 石蕾, 等. 对大数据背景下科学数据开放共享的研究与思考[J]. *中国科技资源导刊*, 2020, 52(1): 1-5.
- [24] 王卷乐, 石蕾, 王玉洁, 等. 科学数据汇聚的模式分析及对我国的发展建议[J]. *地球科学进展*, 2020, 35(8): 839-847.
- [25] Data Cite. Repository Finder[EB/OL]. [2021-01-02]. <https://repositoryfinder.datacite.org>.
- [26] 马军花, 段宗奇, 刘闯. “全球变化科学研究数据出版系统”被美国地球物理学会认定为该学会学术期刊关联原创数据仓储中心全球变化[J]. *数据学报(中英文)*, 2019, 3(3): 305-307, 415-417.

- [27] 国家青藏高原科学数据中心. 国家青藏高原科学数据中心成为国内首个通过 Nature 数据期刊 Scientific Data 认证的数据仓储中心 [EB/OL]. (2020-10-02) [2021-01-02]. <https://data.tpc.ac.cn/zh-hans/news/80e2911c-9a6f-4bf5-a363-ed30fb06a18e>.
- [28] 中国科学院计算机网络中心. 国内首家: ScienceDB 成为 Nature 推荐通用存储库 [EB/OL]. (2020-10-02) [2021-01-02]. [http://www.cnict.cas.cn/xwdt/zhxw/202010/t20201010\\_5714666.html](http://www.cnict.cas.cn/xwdt/zhxw/202010/t20201010_5714666.html).

## Review of the progress of earth data science and sharing in 2020

WANG Juanle<sup>1,2</sup>, WANG Yujie<sup>1</sup>, ZHANG Min<sup>1,2</sup>, JIANG Han<sup>1,3</sup>, HONG Mengmeng<sup>1,4</sup>, LI Shuhan<sup>1,5</sup>

1. Institute of Geographic Science and Natural Resources Research, Chinese Academy of Sciences, State Key Laboratory of Resources and Environmental Information Systems, Beijing 100101, China
2. University of Chinese Academy of Sciences, Beijing 100049, China
3. School of Marine Technology and Geomatics, Jiangsu Ocean University, Lianyungang 222005, China
4. School of Civil and Architectural Engineering, Shandong University of Technology, Zibo 255049, China
5. Institute of Disaster Prevention, Sanhe 065201, China

**Abstract** Earth data science is combination of earth science and data science, and is a frontier cross field of earth science research and a paradigm shift of data-driven scientific discovery. This paper reviews the progress of big earth data research in 2020 from the aspects of earth data science program, scientific data governance, and earth science data sharing. It focuses on the Deep-Time Digital Earth Science Program, Big Earth Data Science Engineering Project and achievements in global data aggregation, authentication, repository, policies, standards, and identification, as well as practices of the earth science data centers, databases, data models, data integration, and data publishing. The paper also points out that it is still an urgent need to pay full attention to the rules of scientific data governance, promote the development of earth science big data programme, strengthen the establishment of earth science data ecology, and improve all aspects of cooperation in the field of earth science data.

**Keywords** earth science; data science; data sharing; data governance; big data ●



(责任编辑 徐丽娇)