面向数据密集型科研的科学数据管理模式研究

支凤稳[1,2,3] 史 洁[1] 郑彦宁[2]
（1.河北大学管理学院，河北保定 071002；2.中国科学技术信息研究所，北京 100038；
3.河北省数字治理与协同治理研究基地，河北保定 071002）

摘要：数据密集型科研范式下，科学研究中产生越来越多的科学数据，科学数据管理变得尤为重要。在梳理国内外相关研究的基础上，剖析数据密集型科研环境下对科学数据管理的新要求，结合数据生命周期理论分析科学数据管理各个阶段任务，构建相应的科学数据管理模式，提出多方主体实践的对策建议，以推动大数据时代的科学数据管理与共享。

关键词：数据密集型科研；科学数据；管理模式；大数据
DOI：10.3772/j.issn.1674−1544.2024.03.007
CSTR：15994.14.issn.1674.1544.2024.03.007
中图分类号：G350
文献标识码：A

Research on Scientific Data Management Model for Data Intensive Research

ZHI Fengwen[1,2,3], SHI Jie[1], ZHENG Yanning[2]
(1.School of management, Hebei University, Baoding 071002; 2.Institute of Scientific and Technical Information of China, Beijing 100038; 3.Collaborative Digital Governance Research Base of Hebei, Baoding 071002)

Abstract: Under the data-intensive research paradigm, more and more scientific data is generated in scientific research, and scientific data management has become particularly important. On the basis of reviewing relevant research at home and abroad, this article analyzes the new requirements of data-intensive scientific research for scientific data management models, analyzes the tasks of each stage of scientific data management based on data life cycle theory, and constructs corresponding scientific data management models. Finally, propose countermeasures and suggestions for multi-party practice, in order to promote scientific data management and sharing in the era of big data.

Keywords: data intensive research, scientific data, danagement model, big data

0 引言

随着大数据时代的到来，数据密集程度的加深，科学研究在经历了实验科学范式、理论科学范式以及计算科学范式后进入数据密集型科学范式。正如图灵奖得主Jim Gray在《第四范式：数据密集型科学发现》中所言，科学研究活动已迈入数据密集型第四范式阶段[1]。吉姆·格雷提出E−Science环境和科学研究第四范式的概念，E−Science为科学研究提供了一种全新的思维与科研

作者简介：支凤稳（1987—），女，河北大学管理学院副教授，中国科学技术信息研究所博士后，研究方向为科学数据、竞争情报；**史洁**（1999—），女，河北大学硕士生，研究方向为科学数据管理与共享；**郑彦宁**（1965—），男，中国科学技术信息研究所研究馆员，研究方向为情报学理论与方法（通信作者）。

基金项目：国家科技基础条件平台中心课题项目“面向数据密集型科研的科学数据管理应用模式与技术研究”（2022WT20）；河北大学雄安新区研究专项“雄安新区数据要素共享机制研究”（2023HXA012）。
收稿日期：2023 年 10 月 2 日。

模式，运用各种工具处理现代科学研究中产生的大量科学数据[2]。科学数据不仅仅是产生于科研活动中，而且还是后续科研活动的基础。数据密集型科研范式就是利用计算机技术和工具采集管理分析数据，挖掘新的知识和研究领域，对科学数据管理至关重要，加强和规范科学数据管理也是提升我国科学研究和科技创新能力的重要方式和手段[3]。

1　相关研究述评

1.1　数据密集型科研范式研究

数据密集型科研范式是一种科研方法论，强调在科学研究中广泛收集、分析和利用大规模数据集的重要性。近年来，国内外许多学者研究数据密集型科研环境所带来的影响。

面对数据密集型科研下数据资源的大规模需求，开展了科学数据管理服务研究。如江波[4]构建了面向数据密集型科研数字图书馆参考咨询服务的模式；朱维乔[1]从科学大数据接收层、存储层、计算层和分析层等构建了科学大数据服务平台模型；邓仲华等[5]基于云计算相关技术构建了数据资源云平台，旨在解决数据密集型科学研究中面临的数据处理问题；顾立平[6]明确了科研模式变革下的数据管理服务各利益方及主要内容等。

针对大数据科研环境压力下科研人员数据素养的研究。如张军[7]发现第四科研范式环境下科研人员数据素养能力不足的问题，并构建科研人员数据素养能力培养框架；凌婉阳[8]提出在大数据科研范式环境下提高科研人员科研数据认知和科研数据能力的建议；Koltay[9]强调了科学工作者数据素养的重要性，并研究了数据密集型科学研究范式下研究人员的预期和现实行为。

针对数据密集型科研环境下科学数据的处理及应用研究。在科学数据处理方面，大数据带来了数据处理、数据存储、数据分析和数据可视化方面的困难。面对这些困难，Fernando等[10]探索了数据密集型应用的可选体系结构，提取数据密集型应用程序的基本需求，将它们转化为易于伸缩的构建块，从而更好地利用高度并行的技术，以解决系统出现的新问题；Saif等[11]研究了基于数据管理和复制的数据密集型云计算系统，提供了优化管理（如存储、复制、过滤等）大数据的方法。在科学数据应用方面，大数据对于企业生产力和学科交叉融合发展都具有极其宝贵的价值，未来的企业发展和科学研究必将聚焦于大数据的探索上。如Chen等[12]阐述了大数据在商业活动、政府管理以及在科学研究中的应用，如在天文学、气象学等学科中需要从大规模的科学数据中探索知识。

1.2　科学数据管理模式研究

对于科学数据管理与共享模式，学者们尚持有不同的观点。但总体可以分为自上而下和自下而上两种模式，自上而下的科学数据中心管理模式鼓励研究人员将数据文件上传到政府或科研机构合作建立的数据平台或数据中心上，以加大资源共享[13]。如清华大学的中国经济社会数据中心[14]建立数据智库为一流大学提供数据申请、处理等服务。自下而上的领域管理模式是指各个学科领域将所产生的科学数据存储在本高校或研究机构所建立的科学数据库中，并进行开放共享。如武汉大学图书馆的高校科学数据共享平台[15]，这个数据库收录社会学、地理学、医学、信息管理学等学科领域所产生的科学数据。国内科学数据管理模式研究主要围绕数据生命周期理论。如储文静等[16]基于科学数据生命周期理论，从联盟架构构建、虚拟工作组组建、管理机制建立3个方面构建科学数据联盟管理模式；储节旺等[17]通过构建数据生命周期模型，进一步构建科学数据管理体系；张迎等[18]在科学数据管理生命周期理论基础上构建了科技文献和科学数据一体化科学数据管理应用模式。

欧美等发达国家和地区不仅在理论上对科学数据管理模式和方法进行研究，还在实践中逐步形成了相对成熟的管理模式。如约翰·霍普金斯大学建立的科学数据服务网站，负责科研数据的归档与共享等，支持研究人员获取共享[19]；美国的ICPSR不仅保管其成员机构提交的科学数据，

而且对外提供数据访问接口，通过获取、开发、存档和传播科学数据，来促进社会科学及相关领域的研究和教学[20]。对比国内外数据共享空间的科学数据管理模式，国外在管理实践、功能设置以及用户服务等方面都比国内完善[21]。相较于欧美国家，我国应构建符合中国实际需求的科学数据管理模式。

相关研究发现，尽管已有学者构建了科学数据管理模式，但很少考虑到大数据时代对科学数据管理模式的要求。鉴于此，本文将从数据生命周期理论出发，分析其在数据密集型环境下的新需求，以此从科学数据管理过程、支撑手段和指导理念、管理与共享机制 3 个方面构建科学数据管理模式，并提出相应的实践对策，以更好地满足数据密集型科学研究的需求，推动大数据时代的科学数据管理与共享。

2 数据密集型科研对科学数据管理模式的新要求

2.1 建立数据关联性

大数据时代，科学数据已成为重要驱动力。为更好地管理大量增长的数据，各国纷纷建立科学数据中心，持续汇聚和整合本国乃至全球科学数据资源，促进科学数据综合利用[22]。现阶段，我国科学数据中心建设虽初具规模，但仍不足以应对数据密集型科研的发展，对数据资源的管理也比较分散。数据中心的建设应与当前数据密集型科研的发展相适应，有效整合数据资源，加强数据中心权威性，在数据驱动创新中发挥重要作用。现在，科学数据不仅仅是资源，更是有价值的资产，因此应利用大数据技术构建科学数据管理与应用模式，在大数据中建立关联性，将潜在的数据价值应用于新的科学研究中，提高科学数据重用能力，使资源的开放利用共享更加有效。

2.2 管理模式要智能化、动态化

在管理过程方面，科学数据管理贯穿整个数据生命周期，应该让数据资源流通起来，形成动态循环，使管理过程更加灵活高效。在管理意识方面，科学数据管理需要投入时间、人力、财力和技术资源。然而，科研机构和团队往往数据管理意识不足，致使大量科学数据流失。大数据时代，科学数据的管理不在于掌握多少数据，而是在于处理数据的能力，将大量的数据转化为有研究意义的科学数据。科研人员应该意识到数据管理是动态化的过程，鼓励科研人员主动参与数据管理，现在更要求利用智能化技术，多手段保障数据安全，维护数据生产者和使用者权益。在管理安全方面，数据密集型环境导致数据泛滥，一系列问题也随之出现，如数据权属、知识产权问题。

2.3 应融合新兴技术

随着互联网技术和通信技术不断提升，各种科研活动中积累的海量数据更新速度快，使得数据存储和处理受到严峻的挑战，再者科学实验和理论研究不断创新，这些创新过程所产生的海量数据也使科学数据爆炸增长。数据密集型科研范式的发展离不开大数据技术的支持，大数据具有数据体量大、数据类型多样、处理速度快、价值密度低[23]等特征，传统的数据管理技术已无法有效地应对现代科研人员科研需求[5]。将大数据技术嵌入数据密集型科学研究中能够提高数据采集、数据存储以及数据处理能力。大数据时代下的数据管理模式更应体现在利用机器学习、人工智能或数据挖掘等技术从大量的科学数据中挖掘潜在知识，并广泛应用在各个领域，最终达到提高数据利用率、推进科学研究的目标。

3 不同生命周期阶段的科学数据管理任务

科学数据管理是科学研究中一个重要的环节，能够帮助科研人员更好地收集分析科学数据，提高研究效率。科学数据管理的定义各有不同，计算机图灵获得者Jim Gray认为数据管理活动涵盖元数据创建、标准制定、数据仓储等广泛的活动[24]。郭佳璟[25]认为科学数据管理主要是对可以用于科学研究的数据或者科学研究中产生的数据进行收集、整理、保存、共享和跟踪。本文认为科学数据管理是指根据数据生命周期的特点利用计算机技术对科学数据进行收集存储、组

织加工与共享应用的过程。针对科学数据管理流程，学者们提出了不同的生命周期模型。储节旺等[17]认为数据生命周期阶段除了数据收集、保存、共享还应包括数据再利用；陈欣等[26]基于国内外46个生命周期模型，从数据创建、数据分析、数据公开3个阶段揭示了社会科学数据从产生到利用的过程；夏义堃等[27]认为应对数据生命周期各阶段进行质量控制；高飞等[28]基于科学数据管理标准、数据汇交与加工、数据长期保存和数据共享服务4个方面，构建农业科学数据生命周期管理模型。

比较一些典型的数据生命周期理论认为数据生命周期主要包括数据收集、数据处理和数据利用等环节。但是针对现在数据密集型科研环境，科学数据管理应该进一步细化明确各个阶段。为此，本文在大数据时代背景下，结合数据密集型科研范式特点与生命周期理论，将科学数据管理分为6个阶段，并且对各个阶段的具体任务进行了分析，如图1所示。

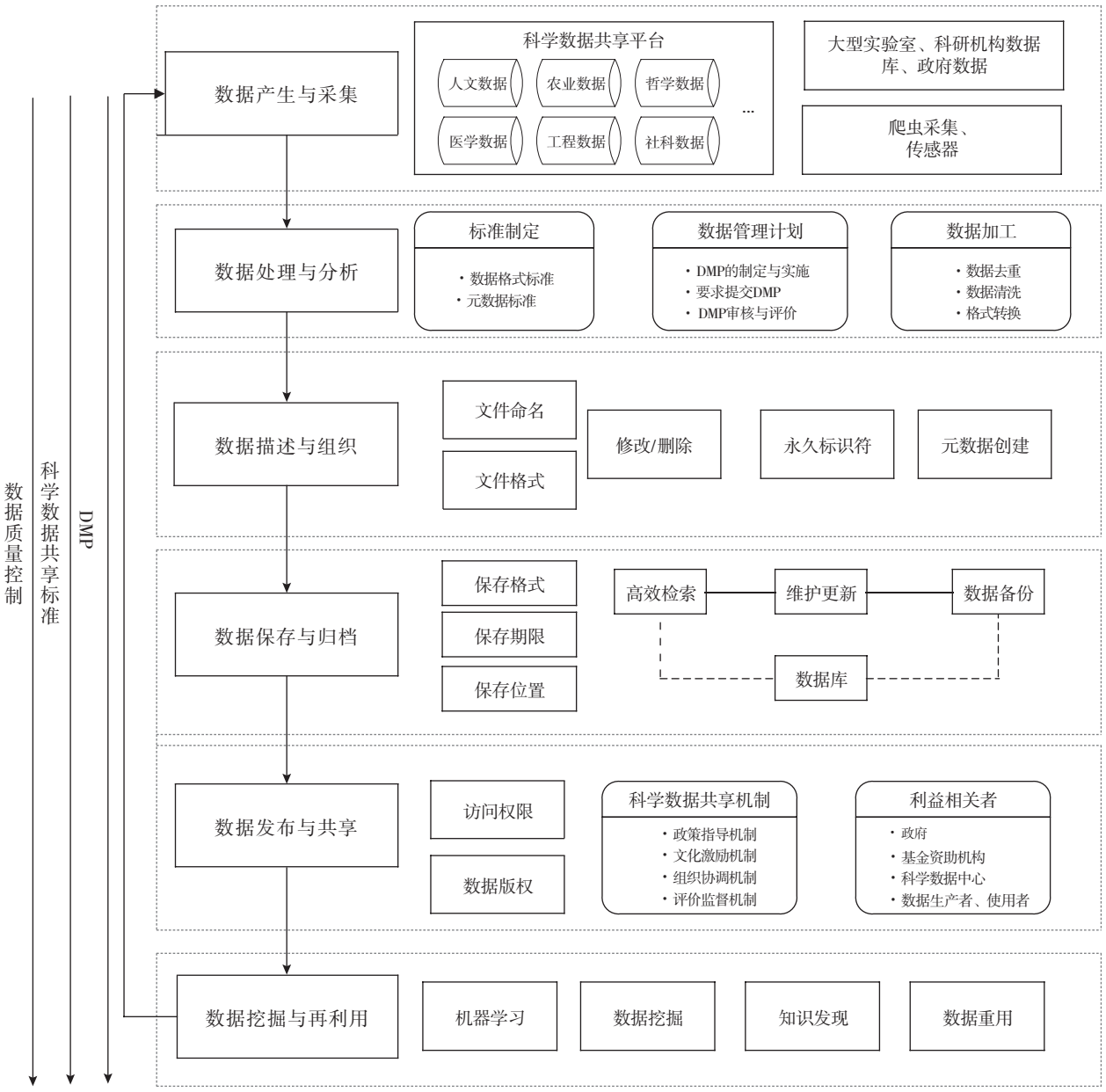

图1　科学数据管理生命周期阶段及任务

3.1 数据产生与采集

科学数据来源广泛，其生产者包括科研人员和科研单位。科研人员可以通过直接获取和间接采集获取数据。前者一般是从科学数据共享平台直接获取各学科的科学数据包括人文社科数据、农业数据、工程数据及医学数据等；后者是通过其他途径来获取公开的或者非公开的科学数据，如利用企业内部的数据库和外部的数据接口等。有些科学数据存储于大型实验室、科研机构数据库、政府数据库中，科研人员可以根据数据标准下载共享这些科学数据。此外，可利用传感器或者爬虫爬取网络数据，爬虫是一种自动化程序，能够模拟人类在网站上浏览，自动抓取内容并存储。大数据时代，爬虫技术日趋成熟，通过大规模数据采集和数据挖掘能够高效快捷地采集数据。

3.2 数据处理与分析

生产或收集的科学数据不能直接保存，还需要经过一定的加工处理和分析。科学数据有结构化、半结构化和非结构化数据，并且涉及多个学科领域，因而制定数据格式以及元数据标准非常重要。关于数据管理计划，国外的科研机构在科学数据管理实践中，都要求提交详细的科学数据管理计划，并且按照DMP对各个阶段进行评估[29]。数据处理就是按照标准转换数据格式，对数据进行清洗，删除不需要的信息，添加所需要的信息，转换成系统可处理易使用的格式。数据清洗能够保证数据分析结果的准确性和可靠性。在大数据科研环境下，人工智能能够识别出有价值的数据，并进行分类，将不准确和重复性的数据进行自动清理，极大程度减少数据冗余。数据密集型科研范式改变了传统的数据分析方式，数据分析过程需要使用一定的分析技术，比如数据挖掘、分析、重组等数据管理技术。科研人员可以通过便利的交互界面进行数据的相关分析，对数据内容进行更深层次的提取与研究，从而提高科学数据利用率，帮助科研人员做出判断。

3.3 数据描述与组织

由于科学数据来源的广泛性和学科交叉性，数据描述与组织是数据管理中重要的一环。数据描述与组织就是将收集到的数据转化为计算机能够处理的形式，并且按合适的元数据标准进行数据组织。对形成的文件命名要具有描述性、唯一性、可读性等，数据的保存格式应符合开放性、通用性等特点[30]。文件格式有图文、视频、音频等，要使用便于存储的方式，还要根据数据之间建立的联系进行数据标识，为后续数据交叉应用奠定基础。

3.4 数据保存与归档

数据存储就是将处理好的数据按照一定的形式存放在物质载体上，以便之后的数据共享和利用，在数据存储过程中要对数据进行分级分类管理，并保证数据的安全性，方便数据重用与查找。在保存格式上，一般都要求采用通用格式进行保存，但个别部门会有专门的数据格式要求。在保存位置上，一些科研工作者会存储于个人设备中，比如计算机、手机、U盘等，但一些在线数据或者大规模数据一般存储于大型机构数据库或第三方数据云存储平台中。数据密集型科研环境对数据存储技术也提出了更高的要求，部分机构为了扩大数据存储空间会引入云计算技术，使数据存储更具扩展性。在保存期限上，各研究机构的要求有所不同。如英国生物技术、生物科学研究理事会和奥地利科学基金会规定在项目结束后至少可以保存 10 年；美国国立卫生研究院规定数据研究项目结束后可以最少保存 3 年[31]。同时，在数据保存阶段，数据备份和维护更新也十分重要，大数据时代实时数据变得更加重要，有必要对数据进行维护更新，及时处理不需要的数据，并补充实时数据，以保证数据的真实性和持久性。

3.5 数据发布与共享

数据发布是数据共享的前提，数据共享是数据发布的直接目的。数据发布与共享阶段要注意数据版权、访问权限以及数据安全等问题，强调隐私保护，在最大限度内共享科学数据。在访问权限方面，部分机构要求作者提供数据可用性声明，对于不可公开的涉及个人隐私的数据或信息

要给予说明，这极大地加强了数据安全性，平衡了各利益方的诉求，引导各方积极推动科学数据的开放共享。建立科学数据共享机制有助于协调各利益方，共同实现科学数据共享。其中，政策指导机制是政府建立相关数据管理政策，在政策方面引导、规范和监督科学数据共享中各利益方的行为；文化激励机制是建立良好的文化氛围，可以强化科研工作者的共享思维和共享意愿；组织协调机制是强调团队成员分工协作，明确利益诉求，资助机构可以将各利益相关者连接起来，建立利益联盟关系；评价监督机制是对数据的管理过程进行评价与监督，评估合作质量，促进科学数据价值最大化。数据共享的本质是互惠互利，而互惠是科研人员科学数据共享的基础条件[32]。共享并非只为了开放，而是让其他研究者利用数据产生更大的价值。

3.6　数据挖掘与再利用

在数据密集型科研范式环境下，科学技术迅速发展，可以利用各种计算机技术进行数据挖掘，发现科学数据潜在价值。数据挖掘是指从大量数据中发现并提取有用信息的过程，当数据被挖掘并转化为有意义的知识后，再利用这些知识，也就是知识发现的过程。要实现数据价值最大化，就需要进行数据再利用。数据再利用意味着新一轮数据生命周期的开始。通过数据挖掘和再利用，数据的适用对象扩大，更大程度地提高科学数据利用率，实现数据价值增值。

4　面向数据密集型科研的科学数据管理模式构建

为使科学数据管理适应大数据时代的发展，本文在借鉴国内外科学数据管理经验，并结合数据生命周期理论的基础上，尝试从管理过程、支撑手段和指导理念、管理与共享机制 3 方面构建数据密集型科研范式下的科学数据管理模式，如图 2 所示。这个模式以科学数据管理过程为核

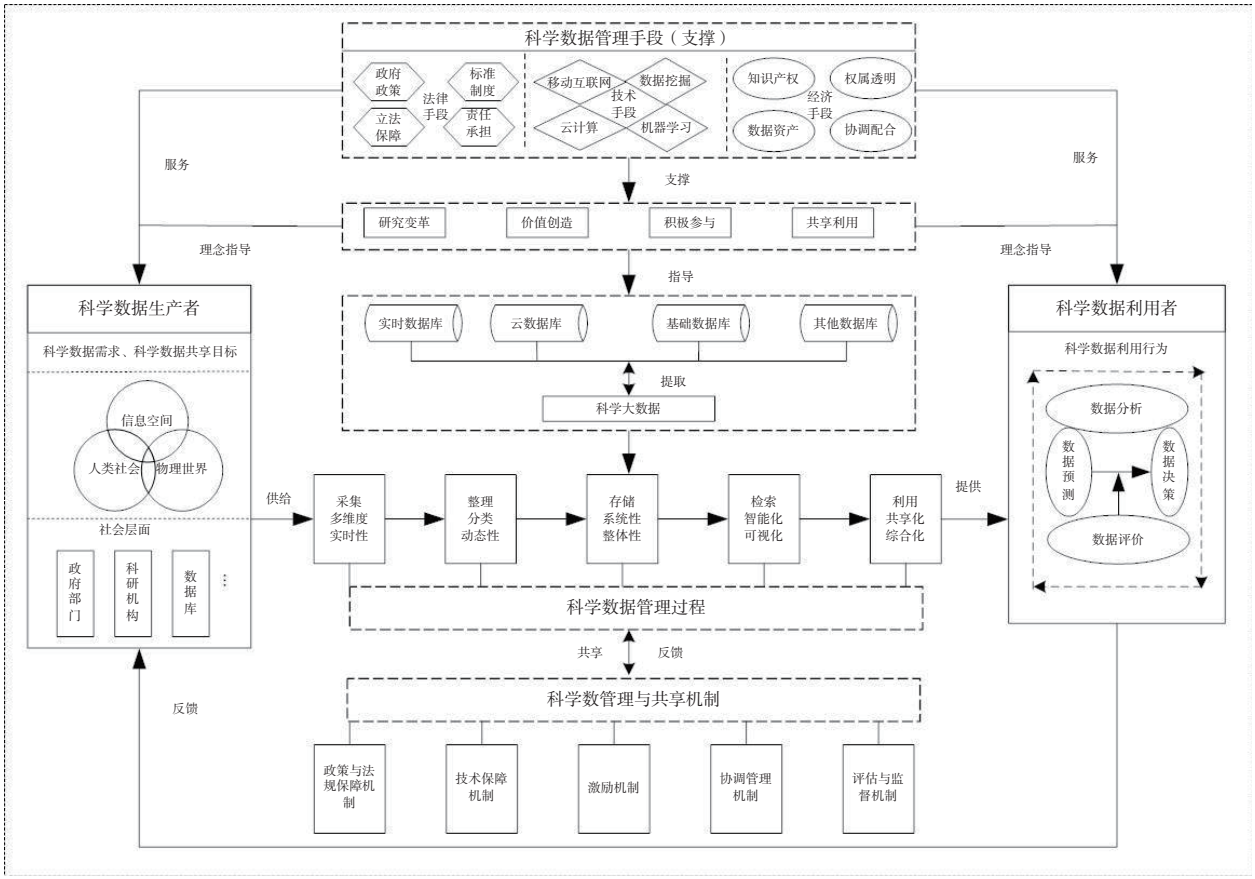

图 2 面向数据密集型科研的科学数据管理模式的理论框架

心，契合数据密集型科研的研究特点，以大数据对科学数据管理的驱动过程为基本模式。

4.1 科学数据管理过程

在大数据时代下的科研活动中，科学数据被赋予了新的使命，无论是实体数据还是异构数据，都需要通过数据管理，组织或可被利用的数据资源[33]。科学数据管理过程，是整个模式的核心。此过程包括科学数据的生产者和使用者、科学数据管理的阶段流程和反馈机制等，主要描述科学数据采集、整理、存储、检索、利用的全生命周期过程，且通过反馈机制实现科学数据管理的有效循环。在数据密集型背景下，科学数据管理过程呈现新特点。采集要注重数据实时性，而且数据获取渠道广泛，要从多个方面多角度获取数据；整理分析要分级分类及动态管理；面对大数据的发展，数据存储的系统性和整体性尤为重要；检索的智能化，建立智能检索系统能够提高查全率和查准率，还可以建立数据检索可视化平台满足用户多样的检索需求；数据利用结合人工智能技术和大数据服务平台实现综合利用，并且重视数据共享。科学数据生产者是模式的逻辑起点，通过科学数据需求，产生科学数据资源，达到科学数据共享的目标。大数据将人类现实空间和信息空间连接成一个整体，科学数据管理与共享离不开人与物的关系，也离不开人与人的关系，科学数据主要来源于人通过大数据获取物理世界和信息空间中产生的科学数据。科学数据利用者是模式的逻辑终点，主要是对共享的数据进行数据管理与利用，如数据分析、数据预测、数据评价和数据决策等，最大化地挖掘科学数据潜在价值。

4.2 科学数据管理的支撑手段和指导理念

科学数据管理模式中需要经济手段、技术手段和法律手段的支撑，同时以价值创造、研究人员积极参与、研究范式改革等理念为指导，推进整个管理过程。数据密集型科研环境下的科学研究具有学科交叉性和开放共享性，不得不考虑数据伦理问题，数据隐私与保护需要法律手段的支撑。在法律手段方面，主要包括科学数据管理制度、立法保障、政府政策、责任承担等，保障科学数据管理过程的顺利实施。在经济手段方面，在数据管理过程中注重知识产权的保护以及数据权属问题包括确定数据权力主体、数据权力内容和数据权力属性，数据资产的协调配合能够为科学数据管理提供一定的经济支持。在科学数据管理过程中最重要的就是技术手段。在技术手段方面，大数据时代的科学数据管理不同于传统的数据管理，人工智能技术影响着数据管理的每个环节，比如云计算技术大大提高了数据存储容量及数据备份、数据恢复功能，机器学习提高了数据处理能力，科学数据管理变得事半功倍。构建数据密集型科研下的科学数据管理模式不仅可以促进数据生产者和使用者的紧密连接和良性循环，而且将大数据技术手段和价值共创的理念真正应用于科学数据管理过程中，使科学数据管理过程更加合理和可靠。

4.3 科学数据管理与共享机制

根据国内外成功经验，运行良好的科学数据共享机制，是保障科学数据管理健康发展的关键[30]。为更好地建立科学数据管理全过程，为科学数据共享服务，有必要建立健全科学数据共享机制。科学数据共享需要在社会多方协作下推进，涉及社会各方利益。因此，要通过制定完善的政策和法律法规来规范数据共享行为，协调和平衡各方利益。如国务院颁发的《科学数据管理办法》对科学数据的共享利用作出规定；部分科技部门制定项目科学数据汇交办法来规范科学数据共享行为，对共享服务进行监督。政策不仅仅涉及共享行为还应包括技术规范，如数据知识产权保护、用户的共享权限、共享数据的保密级别等都需要制定明确的政策与法规。另外，共享技术是科学数据共享的必要条件，没有技术保障是很难实现的，科学数据共享平台需要现代信息技术的支持，如异构数据库技术、区块链技术、集成共享技术等。共享平台要按照统一的共享标准体系搭建稳定的、安全的共享平台，这既是实现科学数据资源共享的核心，也是方便科研人员进行数据共享的基础。在科学数据共享中，提高科

研人员的共享意愿十分重要[34]。共享是需要成本的，最大程度帮助科研人员实现数据共享就要制定激励措施，使不愿意共享的科研人员意识到共享的重要性，强化科研人员的共享思维，激励所有科研人员参与到共享队伍中。为了保障国家利益及个人利益，要建立监督机制对保密数据及用户权限进行监管。数据质量是保证开放共享的基础。高质量的数据有助于研究人员认识到数据共享的价值，并在一定程度上扭转他们的负面态度，从而提高他们共享的意愿，保证数据开放共享的成效[35]。因此，要对数据质量和科学数据共享平台及合作成员进行评估。

5　科学数据管理的对策建议

科学数据管理模式的构建对加深科学数据管理，推动科学数据开放共享具有重要意义，但在大数据时代，科学数据管理的实施还需要多方主体在实践中共同探索。

5.1　政府

政府要完善政策，提供资金保障。科学数据管理工作的开展离不开政策法律的引导和推动[32]，政府应该进一步制定与完善相关政策，如管理策略、管理原则以及相关的数据安全与保护等政策。资本可得性障碍对科学数据共享效果具有反向的影响[36]，建立合理的数据共享激励机制，肯定数据生产的贡献，为其提供必要的经费支持，提升其数据共享行为。

促进新兴技术应用，强化数据管理。首先，要加强对科学数据的管理，为科学数据的广泛获取与开发利用提供支持[37]。其次，要利用好现代信息技术，尤其是新兴的区块链、数据安全加密等技术，加强科学数据管理过程中的数据安全性、数据严密性。最后，科学数据共享正面临着向人文社会科学领域、向微观数据管理、向多学科交叉融合的发展趋势[33]，数据的利用和价值正逐渐受到重视。随着AI、元宇宙以及ChatGPT等新技术的发展，未来将持续出现一些新特征、新模式、新方法，值得继续探索和挖掘。

5.2　科学数据管理平台

科学数据管理平台（一般包括数据中心与数据期刊）是数据存储与传播的重要力量。在新型科研环境下，数据的高速增长对数据存储和传输有了更高的要求。数据类型的多元化，除了结构化数据外，还涌现出更多半结构化甚至非结构化数据，其对数据的管理和集成有着更高的要求，而人工智能、云计算等技术可以满足这些要求，如一些科学数据管理平台可以利用云计算的对象存储技术对大量图片和视频等非结构化数据进行存储和管理。随着人工智能技术的发展，科学数据管理平台不仅是一个简单的数据收集存储设备，而且是具备智能分析和预测能力的数据处理平台。通过人工智能和机器学习技术，平台可以自动识别和分类数据、自动优化存储结构、自动进行数据备份和恢复等，从而提高数据存储和管理效率。平台既需要通过元数据、数据组织与检索等技术手段来保证数据管理工作有序运转，也需要借助组织架构、法律法规等管理措施来提高效率和效用。在数据管理平台方面，平台的安全性受到了使用者的广泛关注[38]。多源数据的异构性可能导致数据被不当使用，影响研究结果的客观性，从而降低信任水平。因此，平台需要严格数据标准和多源数据归档政策，设计通用检索模型，合理处理数据异构问题，并强化有关违反平台规定、威胁数据安全的惩治措施。努力提高平台的权威性，加强科研人员的权益识别和保护，也有助于处理好开放共享和安全保密之间的关系，避免或减少共享的潜在风险，以保护提供者的知识产权，并解决数据安全和隐私问题。

5.3　科研资助机构

科研资助机构可以制定相应的激励机制，以鼓励科研人员将数据汇交到相关的平台统一管理。受资助项目所产出的数据应在相关数据平台进行公开，还应对数据进行中期考核和年度评估，以保证数据质量。从数据采集获取到数据保存汇交，再到数据共享利用，要规范数据管理计划，明确各阶段的主体责任，建立统一的数据标准，以确保数据的可用性和易用性。此外，还需

明确提供者的贡献，并从经济效益和学术声誉等方面给予奖励。确保提供者受益于数据共享，可激励其进行数据共享，营造良好的数据共享氛围。

5.4 科研单位

科研单位应加强对数据共享的宣传，使科研人员和公众认识到数据共享的重要性，并解决产权归属、利益分配、学术评价等关键问题。同时，要定期牵头开展数据共享活动与相关学术交流，鼓励科研人员共享自有数据，并重用他人数据进行科学研究。数据共享与重用既可以降低研究成本，也可以促进科研人员间的交流互动，形成长期互的惠关系。从而使科研人员对数据共享持有更积极态度。此外，还要为本单位科研人员提供保证数据质量的研究条件，并制定相应的控制与评估措施。

加强多方合作，促进学科之间的交流与协作。首先，对外要将科学数据主权牢牢掌握在自己手中，加强国际合作交流，了解国外研究话题和研究热点，拓宽科学数据共享范围；其次，对内注重跨学科、跨机构、跨地域之间的交流与合作，主动汲取其他学科的理论优势和方法技术，不断拓展自身领域的新天地；最后，在保证科学数据主权的基础上，注重在跨国数据流动方面和各机构之间的合作，还要充分调动利益相关者的共享积极性，尤其是本单位研究者的积极性。

5.5 科研人员

科研人员作为科学数据生产、管理与利用的重要主体，应主动提升自身数据素养，培养数据共享与监护意识。在共享活动中，科研人员不仅要确保数据质量，保障数据的真实性、完整性、权威性和有效性，还要积极学习相关的知识和技能，加强对自身数据的监护。此外，还应在明确个人与团队目标之间的差异性和一致性的基础上，关注团队成员之间的数据共享成效，及时总结经验，为今后的科学数据共享提供必要的借鉴。

6 结语

大数据时代，科学数据已经成为科学研究创新和学科交叉融合发展的基础性资源，加强科学数据管理，促进科学数据开放共享，成为社会普遍共识。科学研究第四范式的兴起，使科学数据管理变得尤为重要。为此，本文在梳理国内外相关研究的基础上，分析新研究范式对科学数据管理模式的要求，并结合数据生命周期理论，阐明科学数据管理各阶段的具体任务，构建面向数据密集型科研的科学数据管理模式的理论框架，有望为科学数据管理平台或工具的开发提供借鉴和参考。但本文也有不足之处，尚未深入到管理模式运行过程中的实际问题，未来研究可结合具体应用进一步推进科学数据管理模式实践的发展。

参考文献

[1] 朱维乔.面向数据密集型科研范式的科学大数据服务平台构建研究[J].图书馆学研究, 2017(13): 22−25.

[2] 朗扬琴, 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(2): 92−94.

[3] 苏靖.大数据时代加强科学数据管理的思考与对策[J].中国软科学, 2010(2): 92−94.

[4] 江波.面向数据密集型科研范式的数字图书馆参考咨询服务研究[J].农业图书情报学刊, 2018, 30(9): 161−164.

[5] 邓仲华, 王鹏, 李立睿.面向数据密集型科学研究的数据资源云平台构建[J].图书馆学研究, 2015(10): 42−47.

[6] 顾立平.科研模式变革中的数据管理服务: 实现开放获取、开放数据、开放科学的途径[J].中国图书馆学报, 2018, 44(6): 43−58.

[7] 张军.面向科研第四范式的科研人员数据素养培养研究[J].图书与情报, 2016(2): 133−136.

[8] 凌婉阳.大数据与数据密集型科研范式下的科研人员数据素养研究[J].图书馆, 2018(1): 81−87.

[9] KOLTAY T.Data literacy for researchers and data librarians[J].Journal of librarianship and information science, 2017, 49(1): 3−14.

[10] FERNANDO M,CAMPO R.Parallel architecture exploration for data−intensive applications[D]. Canada:

University of Toronto,2021:1−125.

[11] SAIF U R,KHAN S J,ZOMAYA S A et al.Performance analysis of data intensive cloud systems based on data management and replication:a survey[J]. Distrib parallel databases,2015,34(2):179−215.

[12] CHEN C L P,ZHANG Chunyang.Data−intensive applications,challenges,techniques and technologies: a survey on big data[EB/OL].[2023−10−02]. https://www.sciencedirect.com/science/article/abs/pii/S0020025514000346.

[13] 谢春枝, 燕今伟.国内外高校科学数据管理和机制建设研究[J].图书情报工作, 2013, 57(6): 12−17.

[14] 清华大学经济社会数据中心[EB/OL].[2023−08−20]. http://www.sem.tsinghua.edu.cn/sercent/jjshsjzx.html.

[15] 武汉大学数据共享平台[EB/OL][2023−08−15]. https://whu.metaersp.cn/databaseList.

[16] 储文静, 李书宁.我国科学数据联盟管理模式构建研究[J].图书馆学研究, 2019(14): 51−57.

[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[J].现代情报, 2020, 40(10): 34−42.

[18] 张迎, 张志平, 梁冰.科学数据管理应用模式的研究[J].情报工程, 2017, 3(4): 71−77.

[19] 陈丽君.约翰·霍普金斯大学科学数据管理服务实践与启示[J].现代情报, 2016, 36(4): 110−114.

[20] 李玉灵.美国ICPSR科学数据管理实践对我国科研档案管理的启示[J].档案学刊, 2021(5): 40−47.

[21] 吴雅威, 张向先, 张莉曼.国外数据共享空间的科学数据管理模式解析及其启示[J].情报理论与实践, 2020, 43(7): 186−193.

[22] 王瑞丹, 杨静, 高孟绪, 等.加强和规范我国科学数据管理的思考[J].中国科技资源导刊, 2018, 50(2): 1−15.

[23] 吴林, 吴超, 吴娥.大数据视域下安全信息资源管理模式研究[J].科技管理研究, 2020, 40(9): 156−162.

[24] HEY T.The Fourth Paradigm:Data−Intensive Scientific Discovery[M].New York, USA: Springer,2012:1−284.

[25] 郭佳璟, 樊欣.国外科学数据管理经验及其对我国“双一流”高校图书馆的启示[J].文献与数据学报, 2019(3): 26−37.

[26] 陈欣, 詹建军, 叶春森, 等.基于高校科学数据生命周期的社会科学数据特征研究[J].情报科学, 2021, 39(2): 86−95.

[27] 夏义堃, 管茜.基于生命周期的生命科学数据质量控制体系研究[J].图书与情报, 2021(3): 23−34.

[28] 高飞, 周国民, 满芮.基于生命周期理论的农业科学数据中心化管理模式[J].大数据, 2022, 8(1): 24−36.

[29] 聂云贝, 刘桂锋, 刘琼.数据生态链视角下科学数据生命周期运行过程分析[J].信息资源管理学报, 2021, 11(2): 69−71.

[30] 艾丽丽.大数据时代图书馆科学数据策展管理模式研究[J].图书馆学刊, 2017, 39(8): 12−15.

[31] 马玲.高校科研人员科学数据共享机制研究[J].情报科学, 2021, 39(9): 80−87.

[32] 支凤稳, 张萌.科学数据共享意愿影响因素实证与仿真研究[J].图书情报工作, 2023, 67(13): 111−121.

[33] 江慧慧, 赵丽梅.科学数据共享障碍及消解措施分析[J].图书馆研究, 2022, 52(3): 36−42.

[34] 支凤稳, 张萌, 赵梦凡.双路径视角下科学数据共享行为的影响因素研究[J].信息资源管理学报, 2021(6): 40−50.

[35] ZHI Fengwen, ZHANG Meng, ZHANG Shuaijie, et al. Can social capital and planned behaviour favour an increased willingness to share scientific data? evidence from data originators[J].The electronic library, 2023, 41(4): 456−473.

[36] 华小琴, 司莉, 李亭.我国科学数据共享中障碍因素分析及其启示[J].图书馆工作与研究, 2019, 41(11): 18−26.

[37] 盛小平, 袁圆.国内外科学数据开放共享影响因素研究综述[J].情报理论与实践, 2021, 44(8): 173−179.

[38] 支凤稳, 赵梦凡, 张萌.科学数据共享需求调查与关联挖掘[J].情报科学, 2021, 39(12): 9−16.