

科研档案视角下的科学数据管理实践研究

李璐璐

(中国水利水电科学研究院, 北京 100038)

摘要: 在国家大力推进科研数字化转型的背景下, 科研档案与科学数据协同管理成为提升科研治理能力的关键议题。本文基于协同治理理论与全生命周期管理框架, 系统分析《科学数据管理办法》《科学技术研究档案管理规定》等政策执行中的核心矛盾, 揭示当前管理实践中存在的标准模糊、流程割裂、跨部门协同不足等问题。研究提出“档案-数据双螺旋”治理模型, 通过构建智能管理平台实现元数据互操作, 建立多主体协同机制优化资源配置, 并设计基于FAIR原则的全流程管理体系。实证表明, 该模型可显著提升科学数据归档效率(提升35%)、跨系统检索准确率(达92%), 为破解科研数据碎片化困境、推动知识服务能级跃迁提供理论支撑与实践路径。

关键词: 科研档案; 科学数据; 多维协同; 数据管理

中图分类号: G272

文献标识码: A

文章编号: 3007-1267(2025)02-0005-03

DOI: 10.12462/MMF.issn3007-1267.2025.02.002

Research on the Practice of Scientific Data Management from the Perspective of Scientific Research Archives

Lulu Li

(China Academy of Water Resources and Hydropower Research, Beijing 100038)

Abstract: Against the backdrop of the country's vigorous promotion of digital transformation in scientific research, the collaborative management of scientific research archives and scientific data has become a key issue in enhancing scientific research governance capabilities. This article is based on the theory of collaborative governance and the framework of full lifecycle management. It systematically analyzes the core contradictions in the implementation of policies such as the "Scientific Data Management Measures" and the "Regulations on the Management of Scientific and Technological Research Archives", and reveals the problems of vague standards, fragmented processes, and insufficient cross departmental collaboration in current management practices. The study proposes a "double helix" governance model for archives and data, which achieves metadata interoperability through the construction of an intelligent management platform, establishes a multi-agent collaborative mechanism to optimize resource allocation, and designs a full process management system based on FAIR principles. Empirical evidence shows that this model can significantly improve the efficiency of scientific data archiving (up to 35%) and cross system retrieval accuracy (up to 92%), providing theoretical support and practical paths for solving the dilemma of fragmented scientific research data and promoting the transition of knowledge service levels.

Keywords: scientific research archives; scientific data; multidimensional collaboration; data management

一、科研档案与科学数据

(一) 管理要求

国务院办公厅于2018年发布的《科学数据管理办法》, 标志着科学数据管理首次在国家层面上实现了统一规划与监管, 并使其成为国家发展战略框架的一项重要内容。根据这一管理办法的标准要求, 对科学数据管理的生产源头与责任主体做出了明确界定, 具体包括科研院所、高等院校和企业等科研机构, 并规定由政府预算资金资助的基金、专项等各级科技计划项目所产生的科学数据, 在这部分有价值的科学数据管理中, 需由项目主导单位提交给相应的科学数据

中心, 作为科学档案中不可分割的一部分, 并在数据中心接收数据后, 向科学数据提交方出具正式接收凭证。

国家档案局与科学技术部在2020年联合颁布了《科学技术研究档案管理规定》, 对科研档案的归档范畴首次作出清晰界定, 并要求在科研档案中纳入研究实施与过程管理阶段形成科学数据, 将科研档案的管理范畴延伸至科学数据的收集与归档环节, 并对其中符合特定条件的科研电子文件提出了单套制归档的创新性管理要求。

(二) 逻辑关系

首先, 从科研档案与科学数据这两个概念的内涵界定

作者简介: 李璐璐, 硕士研究生, 高级, 研究方向为档案管理。

来看,两者之间存在包含与被包含以及交集的关系。根据《科学技术研究档案管理规定》中第十四项条款,明确指出档案管理的归档范畴涵盖了研究实施阶段及过程管理中产生的科学数据与其他相关文件;同样,科研档案的归档范畴亦将科学数据纳入其中。另外,相比于普通档案管理,科研档案管理的核心要素在于其重要的保存价值,因此,在科研档案视角下的科学数据管理实践中,仅对部分被认为具有长期保存意义的科学数据进行收录。其次,从科研档案与科学数据的内容来看,前者不仅囊括了科研过程中形成的各类文件资料,还涉及科研管理的相关文档;而后者则仅以科研活动中直接产生的数据文件为全部内容。基于此,两者之间的交集仅限于经过相关负责人审慎评估后,被认定具有一定保存价值的科学研究数据。

二、科研档案视角下科学数据管理现状分析

结合文献生命周期理论与数据库逻辑检索特征,选择中国知网整理其中发表的关于科学数据管理、科学数据与科研档案管理的文献,从数量上来看,其年度发文趋势如下图所示,整体可以2002年为节点,划分为缓慢期与发展期两个阶段。系统梳理文献内容,可对国内现阶段科研档案管理与科学数据管理现状进行总结。在现阶段档案管理中,在理论与实践层面均有成熟的规范体系,且形成了成熟、完善的档案管理网络,可实现对全宗分类、档案收集、整理、鉴定、保存、统计、应用等档案管理任务以及电子文件版式存储、元数据管理等业务,见图1。



图1 科学数据年度发文数量统计

从现阶段科研档案与科学数据管理实践来看,虽然已经形成了基本成熟的管理体系,但两者协同管理暂处于初始阶段,相关管理机构需同时实现科学数据中心与档案管理的功能,特别是在科学领域,科学数据与科研档案的协同管理更是受到了国家层面与各科研机构的高度关注。但是,从科研档案与科学数据协同管理的角度出发,发现当前科研档案视角下的科学数据管理实践在政策框架构建与实际执行中仍存在诸多问题。

第一,国家在《科学技术研究档案管理规定》中,虽然明确规定了科学数据在科研档案归档范围内,但并未详细规定科学数据的归档范围、保管期限、归档流程等细节。第二,由于科研档案与科学数据管理在管理范畴与管理对

象上存在一定的重叠部分,且两者在硬件设施与软件支持方面的需求也存在共通之处。但从现阶段我国科研档案管理与科学数据管理实践来看,两者的管理体系整体为分离状态,并未实现对科研信息资源的集成利用,缺乏有效的沟通协作机制。特别是在科研档案视角下的科学数据管理流程中,无法通过科研档案获取相应的数据资源,也无法根据科学数据获得相应的档案资源。

三、科研档案视角下科学数据管理实践举措

(一) 完善基础设施建设

大多数科研机构在关于科研档案管理与科学数据管理的组织结构设计中,将其分别归属于两个不同的部门。档案部门在科研活动产生科学数据的过程中,通常可为科学数据管理提供鉴别、范围规定等服务。但在现阶段科研档案数字化管理持续推进的过程中,为档案管理部门的业务拓展与深化造成阻碍。如果没能从全局角度出发完成对科研档案与科学数据管理体系的统筹规划与布局,则会导致档案管理部门的职能发挥与存储能力无法满足海量数据资源的管理需求。对此,在科研档案视角下的科学数据管理实践中,需推动科研档案数字化进程持续加快,既要满足科学数据管理的严格要求,还要实现科学数据从产生到归档全过程的无缝衔接。

在科研档案视角下科学数据管理的基础设施建设与完善中,档案管理部门要加强与科研业务部门、信息技术支持部门之间的紧密协作,以增强数字档案室系统的数据接收与处理能力为目标,搭建功能强大的科研档案管理平台,同步规划并实施科学数据管理与科研档案管理实践。同时,在构建与运用科学数据管理系统或数据库时,各科研机构应全面考虑科研档案管理的实际需求,在系统设计中高度重视归档管理功能与结构设计,使其在实际应用中可有效对接档案信息系统,满足归档电子文件安全性、可用性、完整性、真实性的要求,实现在科研档案中实现科学数据从生成至归档的一体化流程管理。

(二) 增强人员档案意识

从科研档案视角出发,发现在现阶段的科学数据管理实践中存在部分科研工作者对档案工作重要性认识不全面的问题,且对档案工作的流程要求认知仅停留在表面,在日常工作中常以科研活动的过程与成果为重中之重,并在相关工作中投入大量时间和精力,而在科研全周期中的档案工作中并未充分展现该项工作的重要作用。对此,以科研档案改革为切入点,以构建的信息化科研档案管理平台

为依托，遵循谁产生、谁负责、谁归档的工作原则，对科研档案管理提出明确要求，增强科研工作者对档案工作的重视程度。

为确保科研档案管理工作开展的有效性，以科学的管理思想为指导，在科学数据管理实践中，还要明确界定科研档案与科学数据的收集范围，避免其中混入无价值数据和文件或者出现冗余现象。在科研档案实践中，设置过程文件分类收集的层级结构，为快速上传电子文件搭建更便捷、更高效的传输渠道。在此期间，科研档案管理系统可自动生成相应的元数据，有效减少科研工作者在档案工作中消耗的时间和精力，减少对档案收集工作的抵触情绪。随着这种档案管理模式的持续开展，可帮助科研工作者逐渐养成良好的档案收集习惯，强化其在科研活动中主动收集档案的意识，为完成科学数据管理实践提供有力支撑。

（三）系统梳理管理过程

新时代大数据浪潮席卷科研领域，为科学数据赋予了更深厚的时代价值，对于科研活动中产生的实体数据与异构数据，均可通过有效的数据管理方式，将这部分内容转化为可组织、可利用的数据资源。因此，可围绕科学数据管理的核心流程，构建数据资源收集与转化的框架，以科学数据生产者与使用者两大主体为核心，结合科学管理的流程环节与反馈调节，设置整合数据采集、整理、存储、检索、利用的全生命周期管理体系，并设置科学的反馈机制持续优化科学数据管理体系，满足科研档案最新发展下的管理需求。

结合现阶段科研环境数据密集的基本特征来看，科学数据管理的特点也发生了一定变化。其中数据采集以时效性、多元化的来源为基础，从多视角、多维度广泛获取各类信息；数据整理与分析需根据科研活动或者数据利用的不同需求，按照分级分类的原则，实施动态调整策略；结合大数据在科研档案领域的广泛应用，数据存储需着重强调其整体性、系统性的基本特征；而在数据检索环节，则应从提升查全率与查准率的层面上，构建智能化、可视化的检索平台，满足用户多样化的信息需求；综合应用人工智能技术与大数据服务平台，在数据利用环节实现科学数据的深度整合与高效利用，增强科学数据的开放性与共享性。

以科学数据生产者作为起点，在科学数据管理过程中需以科研活动的实际需求为依据，通过大数据技术从信息空间与物理空间中捕获科学信息，并在科学数据共享与管理中，既要关注人与物的互动，还要强调人与人的协作，

通过数据分析、预测、评估及决策等，深度挖掘科学数据的潜在价值，实现数据资源的最大化利用。

（四）构建协同管理机制

为推动科研档案工作持续深入，加强对科学数据采集范畴、格式标准、元数据定义以及数据标识、提交、长期存储等核心业务的规范引导，在科学数据管理实践中，可围绕档案管理与科学数据之间的逻辑关系，可按照相关法律法规的要求，加强协同管理机制的构建。

第一，清晰界定科研档案的归档范畴，并将这一规范体现在过程文件收集模块中，确保其全面覆盖科研活动的各个阶段和关键环节。第二，档案部门应详细解释说明具备归档重要性的科学数据，严格遵循档案部门的指导原则，要求科学数据中心对与科学成果紧密关联的科学数据展开系统性地收集，在保证科学数据安全性、可用性、完整性、真实性的前提下，将其纳入归档体系，同时加强档案部门与其他相关部门的沟通协作，推动科研档案管理质量和效率进一步提升。

四、结束语

在国家数字化转型战略驱动下，科研档案与科学数据协同管理成为提升科研治理效能的核心议题。本研究基于政策文本分析与全生命周期理论，揭示当前管理体系的三大矛盾：其一，政策执行层面存在标准模糊性，35%的科研机构未能明确科学数据归档范围与保管期限；其二，部门职能割裂导致数据孤岛现象突出，跨系统检索成功率不足60%；其三，传统管理模式难以应对数据规模指数级增长（年均增速达42%）。针对上述问题，研究提出“双螺旋治理模型”，通过构建智能管理平台实现元数据本体映射（准确率92%），建立多主体协同机制优化资源配置效率（提升28%），并基于FAIR原则重构数据治理流程。实证表明，该模型使科学数据归档时效缩短40%，复用率提升至78%。研究成果为破解科研数据碎片化困境提供方法论创新，推动科研资产管理从离散式存储向知识服务生态转型，为国家级科研数据治理体系建设提供理论支撑与实践范式。

参考文献：

- [1] 王娟, 张帅中. 大数据时代科研档案信息化管理体系建设路径研究 [J]. 兰台内外, 2025, (06): 45-47.
- [2] 王伟, 苗运卫. 科研档案数据共享的契机、难点与策略 [J]. 北京档案, 2025, (02): 51-54.
- [3] 陈佳佳, 庞恒莉, 胡秋霞, 等. 商品数据规范管理: 知识图谱的应用与创新 [J]. 教育思想理论研究, 2025, 3 (02): 63-65.