

科技计划项目科学数据汇交共性机制研究

徐波¹ 王瑞丹¹ 王卷乐² 王玉洁² 王超¹ 高孟绪¹

(1. 国家科技基础条件平台中心, 北京 100038;

2. 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)

摘要: 在大数据时代, 以科学数据为驱动的科学研究已成为新的科研范式, 科学数据日趋成为科技创新的重要产出和动力。科技计划项目是我国支持科技创新活动的重要方式, 也是我国的科学数据产出的重要渠道, 有效开展科技计划项目数据汇交对于规范科研项目管理、持续积累科学数据资源具有重要的意义和价值。本文针对我国科技计划项目科学数据汇交管理的现状和需求, 结合国外科研项目数据汇交管理经验, 辨析我国科技计划项目数据汇交过程中的重要责任主体, 提出科技计划项目科学数据汇交的通用流程及汇交的主要内容, 为研究我国科技计划项目科学数据汇交提出思路和建议, 以期为推动我国科学数据汇交发挥基础性作用。

关键词: 科技计划; 科研项目; 科学数据; 数据汇交; 标准

中图分类号: G311

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2021.01.002

Research on the General Standards of Scientific Data Archiving of the Science and Technology Projects

XU Bo¹, WANG Ruidan¹, WANG Juanle², WANG Yujie², WANG Chao¹, GAO Mengxu¹

(1. National Science and Technology Infrastructure Center, Beijing 100038; 2. State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101)

Abstract: In the era of big data, scientific research driven by scientific data has become a new research paradigm, scientific data has become an important output and driving force for technical innovation. At the same time, science and technology projects are an important form of scientific and technical innovation activities and an important channel of scientific data output in China. Effective data archiving of science and technology projects is of great significance and value for the accumulation of scientific data resources in china. In this paper, aiming at the current situation and needs of scientific data archiving management of Chinese science and technology projects, combined with the experience of foreign scientific data archiving management, analyze the responsibilities of each relevant subject in the process of data archiving of science and technology plan projects, and design a scientific data archiving process based on life cycle, define the content of scientific data archiving, and finally the suggestions were summarized, which will play a fundamental role in promoting the exchange of scientific data in china.

Keywords: science and technology plan, scientific research project, scientific data, data archiving, standard

作者简介: 徐波 (1988—), 男, 国家科技基础条件平台中心助理研究员, 研究方向为科技资源管理; 王瑞丹 (1965—) 女, 国家科技基础条件平台中心研究员, 研究方向为科技资源管理 (通信作者); 王卷乐 (1976—), 男, 中国科学院地理科学与资源研究所研究员, 研究方向为地理信息系统与遥感应用研究; 王玉洁 (1991—), 女, 中国科学院地理科学与资源研究所, 研究方向为地理信息共享; 王超 (1987—), 男, 国家科技基础条件平台中心助理研究员, 研究方向为科技资源管理与共享; 高孟绪 (1982—), 男, 国家科技基础条件平台中心副研究员, 研究方向为科技资源管理与共享。

收稿时间: 2020年7月16日。

0 引言

科学数据逐步发展成为传播速度最快、影响面最宽、开发利用潜力最大的科技资源^[1]。随着大数据时代的到来,数据密集型科学研究已经发展成为新的科研范式,在科技创新活动中越来越重视对科学数据的采集生产、加工整理与开发应用,科学数据开放共享的氛围在科技界也日趋浓厚。科技计划项目是我国政府部门支持科技创新活动的重要途径和手段,代表着我国科技创新的最高水平,也积聚了我国各类高水平科技创新资源。在科技计划项目实施过程中,同步产生海量科学数据资源,是我国科学数据的主要积累途径。因此,规范化地开展科技计划项目科学数据汇交,促进科研数据的持续积累与共享应用,是科技计划项目管理和科学数据资源管理领域长期关注的科研问题。而科技计划管理机构在通过规范数据管理健全科研诚信体系建设方面也存在现实的需求。为此,本文将对我国科技计划项目科学数据汇交管理进行深入研究,并为推动我国科学数据汇交提出建议。

1 国内外科学数据汇交管理的实践

1.1 国外科学数据汇交实践

欧美等发达国家高度重视政府资助科技计划项目数据汇交的系统化、规范化建设,汇交科技计划项目资助机构产生的数据已成为欧美国家科学数据资源建设和持续积累的重要推手。美国支持建设了美国国家航空航天局(NASA)、大气和海洋局(NOAA)以及国家生物技术信息中心(NCBI)、国家科学基金会(NSF)等科研资助机构,采取多种科学数据汇聚的政策和标准,要求项目申请书都必须附带数据共享计划,还对发布和共享数据、保护参与者的隐私和数据机密性、保护知识产权、数据标准、数据共享成本、数据共享计划的监督和评估等进行规定,从需求层面推动了科学数据汇交管理的实施,保障了科学数据管理的可持续发展^[2-3]。这些机构同时下设相应的数据仓储平台,如NSF的Dryad数据仓

储平台和NCBI的GenBank数据中心,对科学数据的保存、发现和复用进行专业化管理。

1.2 国内科学数据汇交实践

近年来,我国各级科技管理部门更加关注科技计划项目数据管理与汇交,在多个领域开展汇交实践。科技部先后出台了《国家重点基础研究发展计划资源环境领域项目数据汇交暂行办法》《科技基础性工作专项项目科学数据汇交管理办法(试行)》等,制定了科学数据汇交的标准规范和工作流程^[4],对相关的科技计划项目数据进行汇交整合与开放共享,明确依托国家科学数据中心开展数据接收、审核与开放共享。中国科学院、国家自然科学基金委员会等部门也通过科研信息化专项等方式支持科研项目科学数据汇交,并制定相关的管理政策制度和标准规范体系,汇交整合了一大批科研项目科学数据资源^[5]。

2 科技计划项目科学数据汇交的意义

2.1 加强科技计划项目管理的重要手段

科技计划项目是我国政府支持科技创新的重要方式,中央和地方政府每年通过科技计划项目形式支持了大量科技创新活动,取得了大批重要的科技创新成果。《科学数据管理办法》《国家重点研发计划项目综合绩效评价工作规范》《关于进一步弘扬科学家精神加强作风和学风建设的实施意见》等文件明确提出先汇交科学数据再进行项目验收的要求。科学数据作为科技计划项目重要成果产出和评价执行情况的重要依据,加强科技计划项目科学数据汇交工作是优化科技计划项目管理的重要工作手段^[6]。

2.2 推动科学数据资源整合的渠道

科技计划项目积累的大量科学数据是我国科技创新活动的重要产出和重要基础,也是我国科学数据的主要积累方式^[5,7]。把分散在各部门、各行业的科技计划项目科学数据有机地组织管理起来,可以有效推动海量科学数据的持续积累、有序收集,最大程度促进数据资源的整合,并且可以有效避免相同数据产生过程中的重复性劳动,建立项目数据汇交机制并规范海量分散数据有效

汇聚重要基地，支撑我国科学数据资源整合。

2.3 促进科学数据集成与应用的重要基础

随着数据密集型科研范式快速发展，科学数据的收集、整理、分析、挖掘与利用在前沿科学领域发挥的作用日益显著，从海量科学数据中提炼有效信息和知识发现的能力在一定程度上决定科学研究水平的高低^[8]。因此，规范科学数据汇交工作、明确汇交内容、建立有效的科技计划项目科学数据汇交质量控制体系是实现数据互联互通、促进交叉融合和开放应用的重要前提。

3 科学数据汇交的机制及其管理

我国科技计划项目种类多样、涉及领域众多、层级差异较大^[9]，产出科学数据种类多、数量大、内容广。通过对我国不同领域、不同层面科技计划项目数据汇交特点进行系统分析，结合欧美等国在科研项目科学数据汇交方面的做法及我国科技计划项目管理和科学数据管理方面的实践经验，应该将以采集生产、加工整理、共享应用为主的科学数据全生命周期管理与以立项、过程管理和验收为主的科研项目管理在各主要阶段相互嵌套，配合开展汇交工作，实现科技计划项目科学数据的良好汇交。如在项目立项阶段同步确定汇交方案，在项目执行阶段完成科学数据的采集与加工，在项目验收阶段完成数据汇交并逐步面向社会提供开放共享。

3.1 框架结构

科研项目科学数据汇交具有共性的技术规律，但目前多为探索性工作，尚未形成统一的体系和机制，也缺乏对汇交工作的规范化指导和推进。通过我国科技项目管理部门和国家科学数据中心的汇交实践，结合科学数据与科技计划项目生命周期管理特点，重点对科技计划项目数据汇交相关责任主体的主要类型及其职责、通用的汇交流程及主要汇交内容进行结构化和体系化，进而总结共性规律形成共性机制和通用技术标准。

基于已有实践和科学数据工作特色，提出标准结构如图 1 所示。

3.2 汇交主体及职责

3.2.1 主要汇交主体

科技计划项目科学数据汇交涉及多方职责，既包括项目资助部门、专业管理机构等承担项目管理的机构，也包括数据中心、数据存储库甚至数据企业等数据管理机构，同时也包括项目依托单位、项目负责人、项目承担团队等项目执行机构和人员。在众多参与科研项目管理和数据管理的机构中，有必要厘清科技计划项目数据汇交的责任机构的主要类型，尤其是要求各类机构的主要职责进行标准化。

(1) 在科技计划项目管理方面，随着国家科技计划管理改革的不断推进，国家重点研发计划等科研项目主要委托项目管理专业机构进行日常管理，也有部分科研项目由资助部门直接管理，承担项目管理具体职责。项目科学数据汇交与项目立项、验收等具体环节紧密相连。因此，科技计划项目科学数据汇交中最重要的主体应为具体承担项目管理的机构，这类机构职责也可以有项目资助部门直接承担。

(2) 在科学数据管理层面，科学数据整理是一项专业性较强的工作，随着科学数据的积累，产生了很多专业从事数据专业化加工整理的机构，如国家科学数据中心、数据存储库，甚至一些专业承担数据管理与应用的企业。这些机构通常具有配套软硬件条件及专业的数据工作队伍，随着科学数据工作的不断推进，依托专业机构承担科学数据管理已经成为科技界认可的做法。

(3) 在项目实施方面，科技计划项目一般由项目承担单位和项目负责人共同管理，并强调项目承担单位承担主体责任。科学数据管理是一项长期的基础性工作，考虑到项目承担单位的工作要求和工作需求以及科研人员的快速流动性和不稳定性，科技计划项目科学数据汇交也应该强调项目承担单位的作用，将项目的牵头法人单位作为科学数据汇交的主体之一^[10]，落实科学数据提交单位责任。

3.2.2 汇交主体的主要职责

(1) 对于科技计划项目管理机构，主要承担

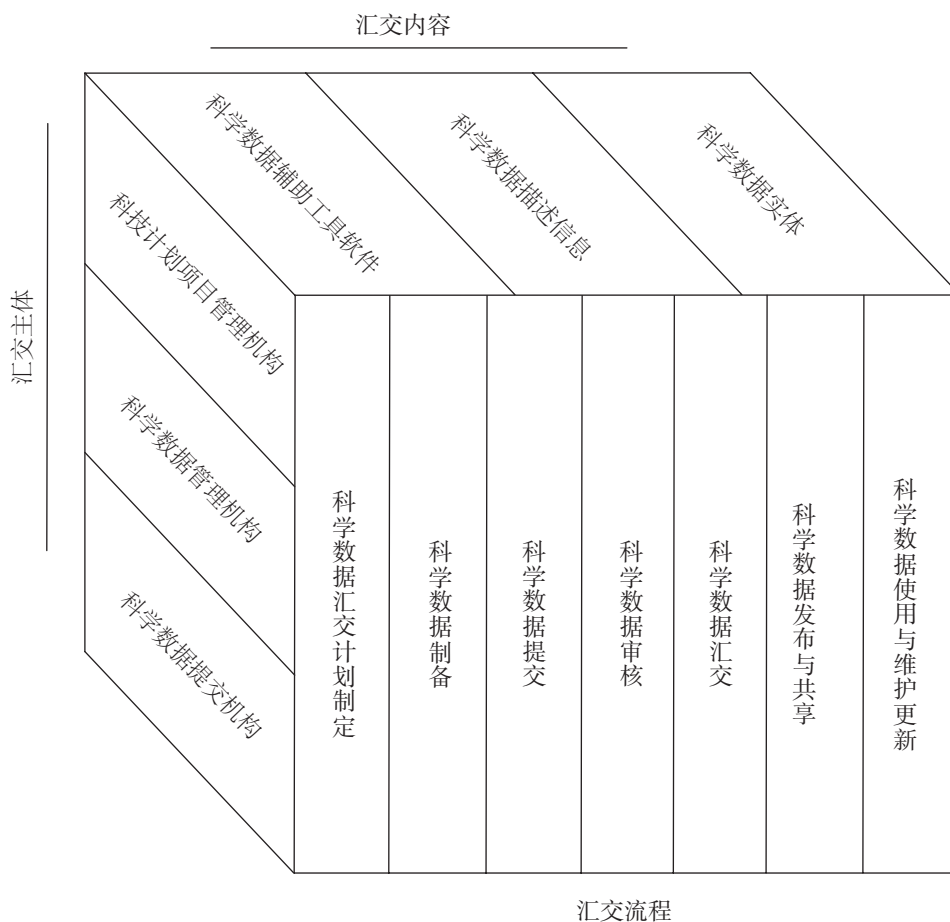


图1 科技计划项目科学数据汇交共性机制框架结构图

科学数据汇交管理职能，负责在项目立项阶段审批科技计划项目数据汇交计划，明确汇交的主要内容和进度安排，在项目实施过程中督促指导项目单位组织开展科学数据收集与整理，并在项目验收阶段组织项目单位将科学数据汇交至相应的科学数据管理机构并对汇交审查报告完成审核。

(2) 对科学数据管理机构，主要承担汇交数据的接收、审查并保存与备份，出具审查报告等职能；负责开展科学数据的分类编目与标识、加工整理和管理维护；发布科学数据及目录，评估科学数据应用情况，提供科学数据共享服务等工作。

(3) 对于科学数据提交单位，主要承担按规定程序和要求向科学数据管理方提交科学数据的责任。在项目申请阶段组织制定科学数据汇交计划，在项目开展阶段对科学数据进行生产、加工和规范整理，并对其质量进行把关，按照科学数

据汇交计划向科学数据管理机构提交科学数据。同时，应该鼓励科学数据提交单位在科学数据发布共享后注重更新和维护数据。

3.3 通用汇交流程

科学数据汇交应与科研项目管理流程紧密结合，在不同阶段由不同的责任主体完成相应的工作任务，在推动科学数据汇交过程中，应重点提出标准化的科学数据汇交通用流程，确定主要环节，并针对汇交流程中的各个环节开展的工作内容进行标准化规定。根据科研项目管理一般规律和科学数据管理生命周期，可将科学数据的汇交分为汇交计划制定、科学数据制备、科学数据提交、科学数据审核、科学数据汇总、科学数据发布与共享和科学数据使用与维护更新等环节，各个环节都要按照相关标准规范开展。具体流程如图2所示。

(1) 在科学数据汇交计划制定环节，科学数

据提交单位应在项目立项阶段按照科技计划项目管理机构的相关规定，编制科学数据汇交计划。在汇交计划编制环节，应对汇交计划的主要内容进行通用的标准化规定，汇交计划应包括计划汇交的科学数据清单、质量控制的说明、所使用的或专门研发的数据软件工具情况、数据开放共享原则以及进行数据汇交的技术方案等内容。不同的学科领域也可按照相应特点对汇交计划的格式进行细化。在科学数据提交单位完成汇交计划后，应由科学数据管理机构进行审查，经科技计划项目管理机构审批后实施，对汇交计划的审查也应按照学科领域特点和科技计划项目管理要求形成标准化的审查规范。

(2) 在科学数据制备环节，科学数据提交单位应按照科学数据汇交计划和相关标准规范，进行规范化的科学数据实体采集与处理，按规定格式形成科学数据的数据元信息。科学数据管理机构也可以根据实际情况参与到科学数据制备过程，对科学数据提交单位的规范化制备提供指导和帮助。

(3) 在科学数据提交环节，科学数据提交单位应按照科学数据汇交计划，对计划汇交的科学

数据进行规范化汇总整理，对科学数据质量按照一定的标准进行自查，编制科学数据质量信息报告。科学数据管理机构按照一定的格式要求接收提交单位提交的科学数据。

(4) 在科学数据审核环节，科学数据管理机构按照科学数据汇交计划和科学数据质量控制体系要求对提交的科学数据进行形式审查，并通过组织同行评议等方式对科学数据质量进行评估。科学数据管理机构应建立审核管理机制，当提交的科学数据通过审核时，向科学数据提交单位出具审查报告作为汇交凭证，并将其提交至科技计划项目管理机构审批；当提交的科学数据没有通过审核时，应及时向科学数据提交单位反馈。科学数据提交单位应及时按照反馈意见进行修改。

(5) 在科学数据汇总环节，科学数据管理机构按照数据通用管理规范对通过审核的科学数据进行分类、编目、标识、保存、加工、整理及管理与维护。科学数据管理机构可根据学科领域特点和科技计划管理要求，制定相应的数据整理与维护标准规范。

(6) 在科学数据发布与共享环节，科学数据管理机构经由科技计划项目管理机构审批对外公布

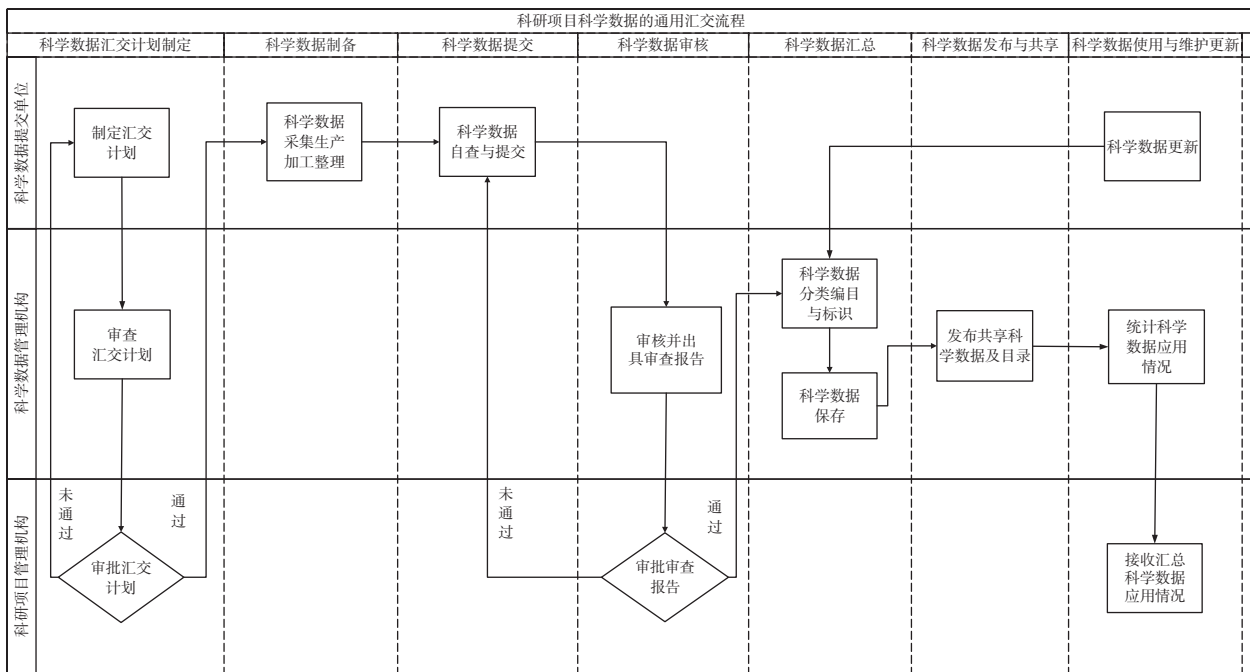


图 2 科技计划项目科学数据的通用汇交流程

科学数据及目录。按照开放为常态、不开放为例外的原则,根据科学数据的开放条件、开放对象和审核程序等,在保障科学数据安全的前提下建立并运行科学数据共享发布系统,提供多种数据快速发现、访问、下载的入口,向用户开放共享。

(7)在科学数据使用与维护更新环节,科学数据管理机构应做好科学数据的管理与维护工作,对科学数据共享应用情况进行统计,并将统计情况反馈至科技计划项目管理机构^[11]。鼓励各科学数据提交单位对所提交的科学数据进行规范化持续更新与归档。科学数据使用者应注重保障科学数据拥有者的合法权益,并在著作、专利、报告等成果中进行体现^[12]。

3.4 汇交内容

不同学科领域科学数据的内容、形式差别明显,数据加工整理所参照的标准规范也不尽相同。为便于汇交数据的整理分析和共享应用,在汇交科学数据实体的同时,还要对相关描述信息及辅助工具软件等内容进行汇交。

科学数据实体是指具体的数据对象,是汇交的主要内容。主要包括在科技计划形成的原始数据及基于原始数据或研究分析数据所形成的完整数据文件或数据库。

科学数据应提供相应的元数据描述信息。元数据的内容、结构和格式等均应符合国家标准的相应规定。各学科领域可根据科技计划项目的实际需要,对元数据进行扩展,并根据实际需要,提供比元数据更详细的数据说明文档及过程资料。

对于用于科学数据处理、加工和分析的专门辅助工具软件等也应提交。提交的科学数据辅助工具软件除了工具软件本身或网络调用接口外,还应包括软件名称、用途、开发工具、运行环境、开发单位、版本号、使用手册等属性信息。对于项目实施中采购的工具软件可不必汇交,但需提供与汇交数据处理相关的工具软件使用说明。

4 结语

本文分析借鉴国外科学数据管理经验,结合

我国科技计划项目科学数据管理的现状和需求,梳理总结我国不同领域、不同层面科技计划项目数据汇交实践,充分总结共性规律,形成科技计划项目科学数据汇交机制,并在国家重点专项等项目数据汇交工作中开始实施。通过规范开展数据汇交,建立工作机制,明确汇交标准,形成长效机制,以期为我国科技计划项目科学数据汇交与管理提供参考。

参考文献

- [1] 国家科技基础条件平台中心.国家科学数据资源发展报告:2018[M].北京:科学技术文献出版社,2019:6.
- [2] 张瑶,吕俊生.国外科研数据管理与共享政策研究综述[J].图书馆理论与实践,2015(11):47-52.
- [3] 王卷乐,祝俊祥,杨雅萍,等.国外科技计划项目数据汇交政策及对我国的启示[J].中国科技资源导刊,2013(2):17-23.
- [4] 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.
- [5] 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):1-5.
- [6] 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.
- [7] 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.
- [8] British Academy. Data management and use: Governance in the 21st century[R]. London: The Royal Society, 2017.
- [9] 胡聪.我国科学数据汇交管理现状、问题及对策研究[J].科技创业月刊,2019,32(7):81-84.
- [10] 秦顺,邢文明.开放·共享·安全:我国科学数据共享进入新时代:对《科学数据管理办法》的解读[J].图书馆,2019(6):36-42.
- [11] 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)[2020-01-08].http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.
- [12] 邢文明,洪芳林,李晓妍.科学数据管理体系的二维视角:《科学数据管理办法》解读[J].图书情报工作,2019,63(23):30-37.