

# 科技计划项目数据管理过程模型

许琦<sup>1,2</sup>, 邹自明<sup>1,2</sup>, 袁雅琴<sup>1,2</sup>, 胡晓彦<sup>1,2</sup>, 佟继周<sup>1,2</sup>, 马文臻<sup>1,2</sup>

1. 中国科学院国家空间科学中心, 北京 100190;
2. 国家空间科学数据中心, 北京 100190

## 摘要

面向未来日益增多的各类科技计划项目和快速增长的科学数据资源, 为了深入加强和规范科学数据管理, 以项目周期为轴线, 融合数据生命周期和数据管护理念, 定义了项目周期内有序开展7项数据活动, 提出了适用于各类科技计划的科技计划项目数据管理过程模型。描述了项目各阶段内依序开展的数据产出分析与管理策划、数据产品定义、产品生产与处理、产品汇交与编目、开放与服务、评价与反馈、维护与保存7项数据活动的内容、输入与产出, 并结合任务书签署、年度/中期检查、综合绩效评价等管理节点给出了数据活动产出控制措施。此外, 提出了项目阶段与管理节点增设、数据活动拆分合并、产出内容与管理角色细化等模型实施建议, 以提高对不同类型科技计划项目的适用性。

## 关键词

科学数据管理; 项目过程管理; 数据管理活动; 数据汇交

中图分类号: G203

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2022002

## *Data management process model for the science and technology programs*

XU Qi<sup>1,2</sup>, ZOU Ziming<sup>1,2</sup>, YUAN Yaqin<sup>1,2</sup>, HU Xiaoyan<sup>1,2</sup>, TONG Jizhou<sup>1,2</sup>, MA Wenzhen<sup>1,2</sup>

1. National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China
2. National Space Science Data Center, Beijing 100190, China

## *Abstract*

To strengthen and standardize scientific data management for the increasing number of science and technology programs and the dramatically growing scientific data resources in the future, seven data activities to be performed in an orderly manner within the program cycle were defined, based on the program cycle as the axis and integrating the data life cycle and data curation concepts. Consequently, a data management process model for science and technology programs applicable to all types of programs was proposed. The contents, the input and output of seven data activities were described, including data output analysis and management planning, data product definition, data product generation and processing, products archiving and cataloging, data open sharing and services, appraisal and comment, maintenance and preservation, which are performed sequentially within each phase of the program. The control measures of data activities output were also given such as document signing and review in accordance with the management nodes of mission statement signing, annual/mid-term inspection, and comprehensive performance evaluation. Furthermore, the suggestions

of model implementation such as adding program phases and management nodes, separating and merging data activities, and elaborating output content and management roles were given to improve the applicability to different types of science and technology programs.

### Key words

scientific data management, program process management, scientific data management activity, data archiving

## 0 引言

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<sup>[1]</sup>。近年来,在《关于改进加强中央财政科研项目和资金管理的若干意见》的指导下,我国科技财政投入快速增长,中央财政科技计划改革取得了决定性进展,形成了布局合理清晰又各具不同管理方式的新五类科技计划(专项、基金等)。根据国家科技管理信息系统公共服务平台内容显示,截至2021年10月,“十三五”期间共部署了68项国家重点研发计划和13项国家科技重大专项,“十四五”国家重点研发计划已启动52个重点专项。随着各级科技计划的有效推进和实施,科学数据将快速积累,科学成果将不断涌现。各类科技计划产生的海量科学数据作为国家资源和人类知识库,特别是以科学探测为目标的项目产出的科学数据具有稀缺性、不可复制性等特点,应进行全面收集、规范管理和长期保存,以便科学团队在未来长时间内可以深入分析挖掘与利用数据背后的知识,促进科学成果产出<sup>[2-3]</sup>。

随着《科学数据管理办法》的发布,科学数据规范化管理的重要性得到了国家层面的高度重视,各类科技计划项目的数据管理细则或要求相继出台,如《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》和《科技计划项目科学数据汇交工作方案(试行)》等,针对正在实施的科技计划项目,确立了项目实施周期内数据汇交、开放共享的管理制度,以综合绩效评价为关键节点,有效管理并促进了项目数据汇交与开放共享。

科学数据的生命周期通常由数据采集生产、汇交保存、开放共享与应用等环节构成,科技计划项目的数据生命周期与项目立项、实施、综合绩效评价和项目后等阶段紧密相关。面向“十四五”时期新的各类科技计划项目,为了进一步促进科技计划项目数据的规范有序汇交,保证数据安全、加强科学数据开放共享与综合利用<sup>[4]</sup>,应将科学数据的管理活动纳入科技计划项目全流程管理,在项目立项、项目实施、综合绩效评价和项目后等各阶段提出数据管理要求或约束,从项目全周期规范数据活动,对数据采集生产、汇交与保存、开放共享等各类数据活动进行过程管理。此外,也应在项目立项前策划各项数据管理活动,将数据管护理念<sup>[5]</sup>融入项目周期内各数据活动,切实提高数据管理质量,保障数据可用性,提高项目数据开放共享水平。

国际上各类数据组织和相关机构通常以构建数据生命周期模型为主要方式,规划指导各类人员做好科学数据管理工作。根据国际卫星对地观测委员会(Committee on Earth Observation

Satellites, CEOS) 信息系统与服务工作小组 (Working Group on Information Systems and Services) 调研报告显示<sup>[6]</sup>, 国外不同机构组织提出了55个数据生命周期模型, 从数据生命周期的视角出发, 描述了数据从产生、收集、描述、存储、发现、分析到再利用的全过程。比如英国数字管护中心 (Digital Curation Center, DCC) 提出了环状管护全生命周期模型 (curation lifecycle model)<sup>[7]</sup>, 国际数据文档倡议联盟组织 (Data Documentation Initiative Alliance, DDI Alliance) 提出了支持社会科学数据管理并具有迭代性的 DDI 组合全生命周期模型 (DDI combined life cycle model)<sup>[8]</sup>, 美国地质调查局 (United States Geological Survey, USGS) 建立了线性的 USGS 科学数据生命周期模型 (USGS science data lifecycle model)<sup>[9]</sup>, 但上述模型并未从项目周期的视角规划数据管理活动。只有美国航空航天局 (National Aeronautics and Space Administration, NASA) 的日地空间物理学科和行星科学数据系统将各类数据活动与项目工作流关联, 在项目周期内定义各项数据活动, 并提供工作指南或参考依据, 极大地促进了项目产出数据的统一规范管理、有序归档和开放共享。NASA 日地空间物理学科为日地空间物理数据环境 (heliophysics data environment, HPDE) 制定了 HPDE 项目数据全生命周期 (heliophysics mission data lifecycle) 模型<sup>[10]</sup>, 该模型将 HPDE 项目周期划分为任务提案、初步设计评审、关键设计评审、任务发射、第一次高级评审、任务扩展、任务运行结束、任务终止、常驻档案馆更新 9 个阶段, 将出具科学提案报告、数据管理计划的准备和签署、数据系统建设部署和测试、数据生产、数据产品优化、数据归档计划的准备和签署等数据管理活动

有序划分到上述 9 个阶段, 以便规范管理。NASA 行星科学数据系统 (planetary data system, PDS) 在项目管理指南中定义了 PDS 项目数据全生命周期模型<sup>[11]</sup>, 该模型以关键评审为节点将项目周期划分为 5 个阶段: 提案、项目研究、定义、设计和研制、任务运行。在模型中详细定义了 PDS 数据归档过程中的数据归档准备、制订数据管理计划、形成数据归档计划、数据产品分发维护、同行评价等管理活动和各里程碑事项<sup>[12]</sup>, 并以文档签署及相关评审活动为节点进行控制管理。

国家空间科学数据中心是 20 个国家科学数据中心之一, 根据《国家科技资源共享服务平台管理办法》, 承担了开展领域内重要科技资源的收集、整理、保存任务, 与领域相关的科技计划项目实施所形成的科技资源的汇交、整理和保存任务, 还开展了科技资源的社会共享、面向科技创新活动及需求提供公共服务和定制服务等任务<sup>[13]</sup>。在推动空间科学领域数据资源的有序汇交、规范管理和开放共享, 保障科学数据安全的同时, 国家空间科学数据中心总结中国科学院空间科学战略性先导科技专项、国家重大科技基础设施项目“子午工程”等项目科学数据管理经验, 以科技计划项目实施周期为轴线, 建立科技计划项目数据管理过程模型, 在周期内的各阶段规划数据生产、处理加工、汇交与保存、共享与服务等各类数据管理活动, 明确数据管理对象与内容, 提出数据管理共性要求。

## 1 科技计划项目数据管理过程

### 1.1 模型概述

科技计划项目数据管理过程模型的管

理对象为各级各类科技计划项目周期及前后阶段内开展的各项数据活动。模型由项目阶段与管理节点、数据活动、项目数据管理角色、产出文档4个要素组成,如图1所示。项目阶段可划分为项目立项、项目实施、项目后3个主要阶段,并设立任务书签署、年度/中期检查、项目综合绩效评价3个管理节点。依据项目周期和项目数据的生命周期,定义了数据产出分析与管理策划、数据产品定义、产品生产与处理、产品汇交与编目、开放与服务、评价与反馈、维护与保存7项数据活动,以及《项目数据管理计划》《数据产品定义文件》《项目数据

汇交计划》及《科技计划项目汇交凭证》等产出文档。将项目数据管理相关人员划分为项目实施方、项目管理方以及科学数据中心,并明确职责,目的在于保证科学数据的真实性、完整性、标准性、可用性和科学有效性,最大限度地满足科研工作者对科学数据的应用需求。

## 1.2 项目立项阶段

在项目立项阶段,主要开展数据产出分析与管理策划活动。项目实施方以项目方案论证报告或其他立项依据相关材料为

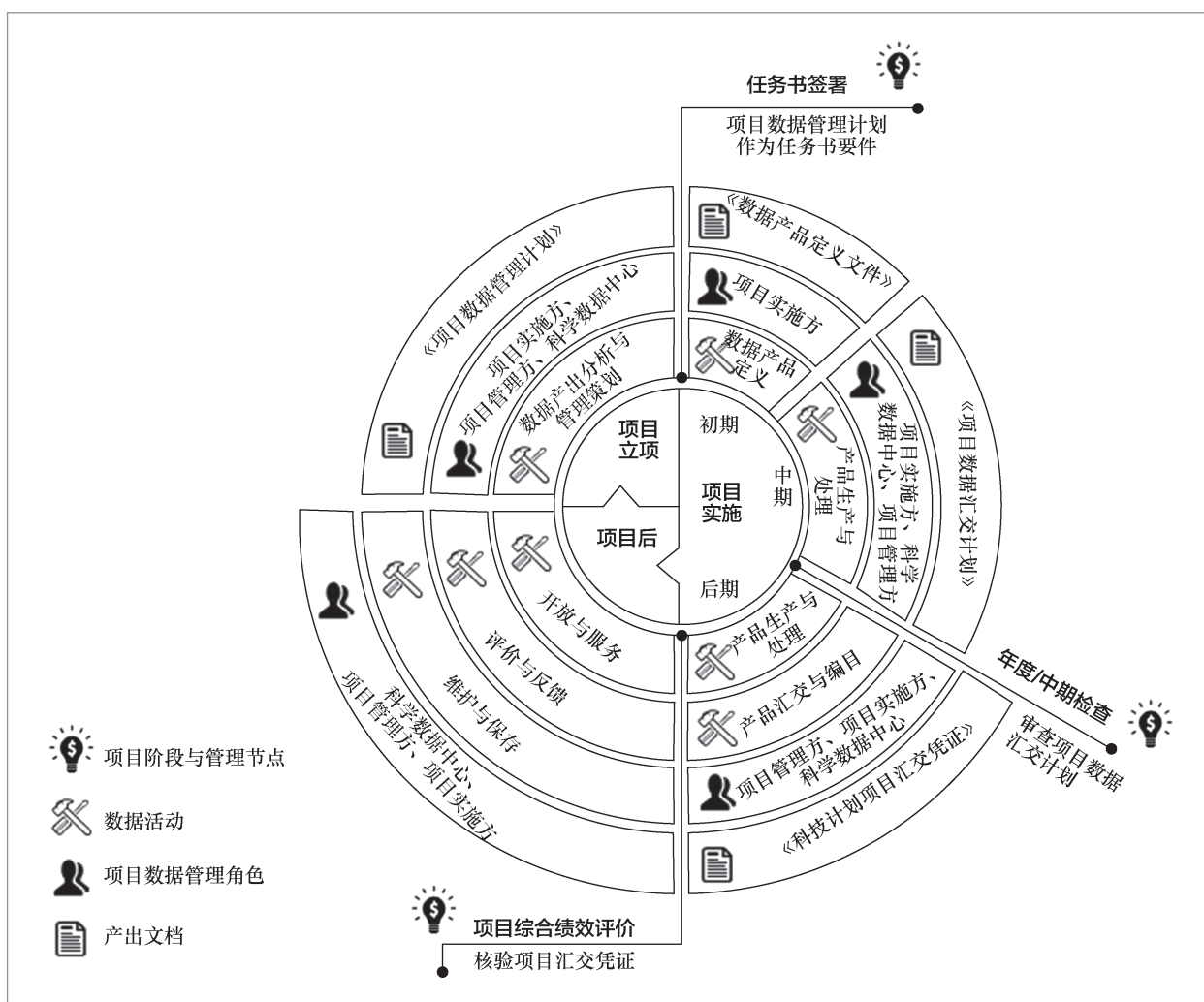


图1 科技计划项目数据管理过程模型

输入,组织开展数据产出分析,明确佐证或支持科学研究目标的数据需求,对项目数据产出的种类、体量以及数据产品预期的质量、应用效果进行需求分析。同时,项目实施方需开展项目数据科学管理策划,科学数据中心需配合参与策划,以确定项目实施方内部数据管理的主要角色及职责分工,确定项目数据采集生产及处理流程,明确项目数据产品设计、采集生产、处理加工、传输、保存等需求,明确数据汇交、共享等方面的管理要求,形成《项目数据管理计划》,并作为项目任务书要件之一。项目管理方应在对《项目数据管理计划》的可行性进行审查后签署任务书。

### 1.3 项目实施阶段

在项目实施阶段,前期项目实施方以项目管理计划为输入,开展数据产品定义活动,定义项目数据产品级别和类别、明确各级各类数据产品的组织、格式、命名和内容,形成《数据产品定义文件》,文件内容应包括数据分级定义与数据产品格式说明等内容,并对数据产品定义文件进行评审。在实施阶段中期,项目实施方以数据产品定义文件为输入,开展数据产品生产与处理活动,包括数据采集生产、数据处理、标定和反演分析以及质量控制,生成各级、各类数据产品。项目实施方应组织开展数据产品同行评价,邀请科学数据中心和相关数据评估专家参与,对项目数据产品的真实性、完整性、标准性、可用性和科学有效性进行客观性评价,形成《数据产品同行评价报告》。此外,项目实施方应开展《项目数据汇交计划》编制工作,根据《科技计划形成的科学数据汇交 技术与管理规范》(GB/T 39912—2021)<sup>[14]</sup>的要求,汇交的主要内容应包括但不限于科学

数据实体、科学数据描述信息及科学数据辅助工具软件等。各类科技计划项目有单独政策要求的,应按相关政策执行。在“开放为常态,不开放为例外”的原则下,可在《项目数据汇交计划》中明确项目数据开放共享相关条件,也可形成单独的《项目数据共享实施细则》文档,用于明确项目各级各类科学数据的保护期限、开放条件、开放对象和审核程序。此外,各类科技计划项目可按需设立数据保护期,但保护期不宜过久,涉密项目数据除外。按照GB/T 39912—2021的要求,《项目数据汇交计划》应由科学数据中心审查,项目管理方审批后实施,因此《项目数据汇交计划》的审批可纳入项目年度/中期检查管理节点,作为必要的检查内容之一。在项目实施后期,项目实施方和科学数据中心共同开展数据产品汇交与编目活动,项目实施方以《项目数据汇交计划》为输入,启动数据产品汇交工作。接收数据产品的科学数据中心以《项目数据产品汇交计划》和相关的《数据产品同行评价报告》为输入,对汇交数据产品进行规范性审查,并反馈问题。科学数据中心应开展产品整编工作,按照相关标准规范对数据产品进行规范性整理与编目,形成便于使用的数据库或数据集,编制数据资源目录,对数据进行分级分类管理。在项目综合绩效评价时,科学数据中心据实出具数据汇交凭证,必要时对汇交凭证进行解读。

### 1.4 项目后阶段

在项目后阶段,科学数据中心同时开展数据维护与保存活动、数据开放与服务活动、数据评价与反馈活动,项目实施方配合参与。数据维护与保存活动以《项目数据管理计划》《项目数据汇交计划》等文件为输入,科学数据中心对汇交的数

据产品进行分级存储、备份和长期维护,按需对数据产品进行版本更新、维护与安全保存。在数据开放与服务活动中,依据《科学数据管理办法》要求<sup>[8]</sup>,除国家法律法规特殊规定的外,科学数据中心应公布科学数据开放目录,并将项目数据开放目录及时接入国家数据共享交换平台,通过在线下载、离线共享或定制服务等方式面向社会和相关部门开放共享。此外,科学数据中心应提供应用服务支持,包括但不限于数据出版与引用、数据确权与知识产权保护等,并对应用情况进行跟踪统计。在数据评价与反馈活动中,科学数据中心组织数据应用情况评价,并将数据应用服务评价情况反馈至项目管理方与实施方。

## 2 实施建议

科技计划项目数据管理过程模型旨在为项目实施方、项目管理方、科学数据中心的数据管理工作提供指南及参考依据,包括为项目管理方在立项阶段的项目数据产出需求和目标评审、项目数据管理计划的评审提供依据,为项目综合绩效评价时的成果评价提供依据;为项目实施方开展规范化的全过程数据管理工作提供指导,包括数据产出规划、数据产品设计、数据产品质量控制、汇交管理和数据成果归档管理等;为科学数据中心制订和审查规范化的数据汇交和共享计划及管理工作提供依据,包括数据汇交、分类分级、规范化整理、数据标识、知识产权管理、数据共享方式、数据应用评价等。

科技计划项目数据管理过程模型满足各类科技计划项目数据的过程管理需求,如国家重点研发计划、科技基础性专项等项目。在模型具体应用实施方面,科技计

划项目数据管理模型的项目阶段与管理节点、数据活动、项目数据管理角色、产出文档4个要素均具有可扩展性,可根据实施建议(见表1)进行适用性改进,形成具体项目和具体领域的的数据管理模型。具体如下。

- 项目阶段与管理节点要素可根据工程技术、基础研究、应用研究、试验开发等不同类型的项目进行扩展。如针对具有建设、运行阶段的工程类项目,模型的项目实施阶段可拆分为建设、试运行、长期运行等阶段,并根据阶段增设相应的管理节点,将项目管理方检查或数据管理过程中的关键性技术评审作为控制活动,如数据采集设备或数据采集处理系统的交付验收测试等。

- “数据产出分析与策划—数据产品定义—产品生产与处理—产品汇交与编目—开放与服务—评价与反馈—维护与保存”7个数据活动组成了完整的项目数据流,贯穿数据全生命周期,可依据不同项目阶段和特点进行拆分合并,如将开放与服务、评价与反馈、维护与保存合并为数据应用活动。

- 模型以产出文档约束相关数据活动,并将产出文档作为下一项活动开展的主要输入,如项目数据产品级别较复杂、产品种类及格式较多、数据生产方较多,可将产出文档细化形成各级、各类、各方的数据管理文档,可按照数据产品级别形成各级产品定义文件;针对具体某一项活动的管理,可依据行业惯例、领域数据资源管理特殊需求等形成单独的项目数据发布共享细则、数据保存计划等,也可将相关内容归纳到项目数据汇交计划。

- 模型中定义了3类通用型的数据管理角色,各类科技计划项目可根据数据管理办法/政策要求,按照项目组织管理结构进行角色细化和拆分,如项目实施方设定

表1 模型实施建议

模型要素	项目阶段与管理节点	数据活动	产出文档	项目数据管理角色	
实施建议	根据项目类型（工程技术、基础研究、应用研究及试验开发）将项目阶段细化，并增设相关评审活动作为管理节点	依据项目阶段和项目类型特点进行拆分合并	某项产出文档可根据数据产品级别、种类、数据生产方等因素细化成多份文档	可根据项目组织管理结构进行细化角色及职责分工	
模型基本框架	项目立项阶段	任务书签署	数据产出分析与管理策划	《项目数据管理计划》	项目管理方 项目实施方 科学数据中心
	项目实施阶段	/	数据产品定义	《数据产品定义文件》	
		年度/中期检查	产品生产与处理	《数据产品同行评价报告》	
		综合绩效评价	产品汇交与整编	《项目数据汇交计划》 《科技计划项目汇交凭证》	
项目后阶段	/	维护与保存 开放与服务 评价与反馈	/		

内部的项目总体、首席科学家、建设运行团队、项目数据中心等，并细化相应工作的数据管理职责。

### 3 结束语

在我国科技投入强度不断增长和科学数据快速积累的趋势下，科学数据汇交与保存、开放共享等方面的规范化管理需进一步强化和加深。为了满足各类科技计划项目数据的过程管理需求，特别是“十四五”时期国家重点研发计划等科技计划项目，本文提出了科技计划项目管理过程模型，旨在从项目立项阶段开始指导项目实施方、项目管理方、科学数据中心全面参与项目全周期内的数据管理活动。通过在项目各阶段根据任务书签署、年度/中期检查、综合绩效评价等节点，将《项目数据管理计划》《项目数据汇交计划》和《科技计划项目数据汇交凭证》等产出文档作为相关活动的控制管理措施，通过《数据产品定义文件》和《数据产品同行评价报告》等评审性活动的产出文档保障项目数据质量控

制，推动项目数据管理工作规范有序开展，保障数据质量。当前，科技计划项目数据管理过程模型正在通过全国科技平台标准化技术委员会（TC486）申请国家推荐性标准研制任务，以期推广扩大应用效益，推动科学数据规范管理、有序汇交、开放共享，助力良性的数据应用生态系统的形成。

### 参考文献:

- [1] 徐波, 王瑞丹, 王卷乐, 等. 科技计划项目科学数据汇交共性机制研究[J]. 中国科技资源导刊, 2021, 53(1): 9-14.  
XU B, WANG R D, WANG J L, et al. Research on the general standards of scientific data archiving of the science and technology projects[J]. China Science & Technology Resources Review, 2021, 53(1): 9-14.
- [2] 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, 26.  
WANG R D, GAO M X, SHI L, et al. Research and thoughts on the opening and sharing of scientific data under

- background of big data[J]. China Science & Technology Resources Review, 2020, 52(1): 1-5, 26.
- [3] 邹自明, 胡晓彦, 熊森林. 空间科学大数据的机遇与挑战[J]. 中国科学院院刊, 2018, 33(8): 877-883.  
ZOU Z M, HU X Y, XIONG S L. Challenges and opportunities of big data in space science[J]. Bulletin of Chinese Academy of Sciences, 2018, 33(8): 877-883.
- [4] 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.  
WANG R D, YANG J, GAO M X, et al. Thoughts on strengthening and standardizing the management of scientific data in China[J]. China Science & Technology Resources Review, 2018, 50(2): 1-5.
- [5] 于明鹤, 聂铁铮, 李国良. 数据管护技术及应用[J]. 大数据, 2019, 5(6): 1-17.  
YU M H, NIE T Z, LI G L. Data curation technologies and applications[J]. Big Data Research, 2019, 5(6): 1-17.
- [6] SINAEEPOURFARD A, MASIP-BRUIN X, GARCIA J, et al. A survey on data lifecycle models: discussions toward the 6vs challenges[R]. 2015.
- [7] HIGGINS S. Draft DCC curation lifecycle model[J]. International Journal of Digital Curation, 2008, 2(2): 82-87.
- [8] VARDIGAN M, HEUS P, THOMAS W. Data documentation initiative: toward a standard for the social sciences[J]. International Journal of Digital Curation, 2008, 3(1): 107-113.
- [9] FAUNDEEJL, BURLEYTE, CARLINOJ A, et al. The United States geological survey science data lifecycle mode[R]. 2013.
- [10] CLARKE S W. NASA heliophysics science data management policy[Z]. 2016.
- [11] Planetary Data System. Planetary data system mission proposer's archiving guide(Version 5)[Z]. 2019.
- [12] 郑岩, 邹自明, 佟继周, 等. 行星科学数据系统(PDS)标准规范的研究[J]. 科研信息化技术与应用, 2009(1): 42-49.  
ZHENG Y, ZOU Z M, TONG J Z, et al. Research on the planetary data system(PDS)[J]. E-Science Technology & Application, 2009(1): 42-49.
- [13] 国家科技基础条件平台中心. 国家科学数据资源发展报告-2019[M]. 北京: 科学技术文献出版社, 2020: 15-16.  
National Science and Technology Infrastructure Platform Center, Ministry of Science and Technology. National science data resources development report 2019[M]. Beijing: Scientific and Technical Documentation Press, 2020: 15-16.
- [14] 国家市场监督管理总局, 国家标准化管理委员会. 科技计划形成的科学数据汇交 技术与管理规范: GB/T 39912—2021[S]. 北京: 中国标准出版社, 2021.  
State Administration for Market Regulation, Standardization Administration of the People's Republic of China. Science and technology program-generated scientific data archiving-technical and management specification: GB/T 39912—2021[S]. Beijing: Standards Press of China, 2021.

## 作者简介



许琦 (1990- ), 女, 中国科学院国家空间科学中心助理研究员, 主要研究方向为科学数据管理、科学数据标准研究。



邹自明 (1971- ), 男, 博士, 中国科学院国家空间科学中心研究员、副主任, 主要研究方向为空间科学信息学。



袁雅琴 (1988- ), 女, 中国科学院国家空间科学中心助理研究员, 主要研究方向为科学数据标识与出版。



胡晓彦 (1987- ), 女, 中国科学院国家空间科学中心副研究员, 主要研究方向为空间科学数据管理与应用。



佟继周 (1976- ), 女, 中国科学院国家空间科学中心研究员、运控部副主任, 主要研究方向为空间科学大数据应用技术、空间科学数据标准研究。



马文臻 (1982- ), 女, 博士, 中国科学院国家空间科学中心副研究员, 主要研究方向为空间科学大数据处理、卫星地面系统技术。

收稿日期: 2021-10-14

通信作者: 邹自明, mzou@nssc.ac.cn

基金项目: 中国科学院“十四五”网络安全和信息化专项资助 (No.WX145XQ07-06)

Foundation Item: The 14th Five-Year Network Security and Informatization Plan of Chinese Academy of Sciences (No.WX145XQ07-06)