

基于科学数据管理流程的科研机构 职责分析

黄源 施栩婕 李晨英
(中国农业大学图书馆, 北京 100193)

摘要: 2018年3月国务院发布《科学数据管理办法》, 强调科研机构作为科学数据管理的责任主体, 应部署本机构科学数据的管理与服务体系。科研资助机构数据政策为科研机构明确了管理科学数据的过程中需要完成的任务与目标, 可以有效推进科研机构的科学数据管理实践。按照科学数据管理流程的核心环节, 系统分析并梳理国际科研资助机构的数据政策, 将政策要求转化为科研机构在数据管理核心环节的职责与任务, 以为我国科研机构有效部署科学数据管理工作提供参考。

关键词: 科学数据; 管理职责; 科研机构; 数据政策

中图分类号: G250

DOI: 10.3772/j.issn.1673-2286.2020.01.003

科学数据既是科学研究成果的重要组成部分, 又是激发新的科研创造所必需的原始材料, 更是数据驱动研究范式下促进科学发展的重要因素。随着开放科学运动的发展, 开放数据和科学数据管理在世界范围被高度重视^[1]。2018年3月17日国务院印发《科学数据管理办法》(以下简称《办法》)^[2], 强调高等院校、科研院所等法人单位是科学数据管理的责任主体, 在本机构部署科学数据管理与服务体系是我国科研机构的重要责任和迫切任务^[3]。截至2019年6月, 2019 US.News全球大学排名前100的大学中有87所大学已经开展了不同程度的数据管理工作, 但国内仅北京大学、复旦大学、武汉大学等少数高校开展了科学数据管理工作^[4], 科学数据管理与服务体系构建仍处于摸索阶段。为贯彻落实《办法》, 进一步提升我国科学数据管理与共享水平, 科研机构在其中发挥着重要作用, 其作用涉及数据收集、加工、保存等数据管理各阶段^[2], 因此, 有必要基于科学数据管理流程深入分析并归纳总结科研机构在科学数据管理过程中的职责, 以为国内科研机构开展和部署科学数据管理与服务体系提供参考。

1 科研机构与科研资助机构在科学数据管理中的关系

科研机构是高等院校、科研院所等法人单位的统称, 既是科学数据的主要产生地, 也是科学数据的重要使用场所^[5]。《办法》将科研机构定义为科学数据管理的责任主体, 对本机构科研人员产生的科学数据负有管理义务。科技计划(专项、基金等)项目产生的科学数据由承担项目的科研机构进行统一管理并向资助机构汇交, 强调科研机构应贯彻落实资助机构对科学数据管理的要求^[2]。

同时, 国内外学者研究表明, 科研资助机构数据政策是促进科学数据管理的重要因素, 指导科研机构的科学数据管理实践进展是科研机构管理科学数据的过程中需要完成的任务与目标。崔雁^[6]从数据存储和政策内容两个维度分析国外科研资助机构的数据政策, 认为科研资助机构作为科研项目的主要资助主体, 在引导科研机构开展科学数据管理工作方面起着至关重要的作用。Jones^[7]提出科研机构应积极应对资助机构要求, 应构建基础设施、增强研究评估并创建数据管理

最佳实践。陈大庆^[8]认为科研机构应遵守资助机构政策要求,辅助科研人员制订科学数据管理计划并进行数字保存。张瑶等^[9]在分析国外资助机构政策内容的基础上,认为科研资助机构的科学数据政策不仅指导科研机构开展科学数据管理工作,还总结科研机构的实施经验。Pink^[10]有针对性地调研了英国巴斯大学为满足其主要资助机构,英国工程与物理研究理事会(UK Engineering and Physical Sciences Research Council, EPSRC)的科学数据管理政策要求而采取的措施,认为科研机构需完成的科学数据管理基础任务即为科研资助机构的政策要求。黄国斌等^[11]认为科研机构的科学数据管理不仅应重视资助机构政策中的共享要求,而且应满足数据管理过程中每一个阶段的要求。因此,本研究主要基于国际科研资助机构科学数据政策,系统地解析科研机构贯穿数据生命周期的管理职责。

2 研究思路与方法

以科学数据生命周期理论为基础,对比分析不同生命周期模型划分的管理阶段^[12],发现各模型中均包含数据管理流程的5个核心环节,即数据计划、数据收集、数据加工、数据长期保存、数据共享利用。本研究基于这5个核心环节,深入剖析科研资助机构的数据政策在每一环节中的客观要求(或目标),然后针对每一环节目标系统地解析科研机构职责。为此,首先利用资助者开放获取政策数据库Sherpa Juliet^[13]调研科研资助机构的数据政策,截至2019年6月, SHERPA登记了26个国家与地区的共计162个科研资助机构的开放获取政策。其中已有64个科研资助机构对科学数据提出不同程度的要求,提出鼓励性建议的有22个机构,制定强制性政策的有42个机构。其国家分布如图1所示。

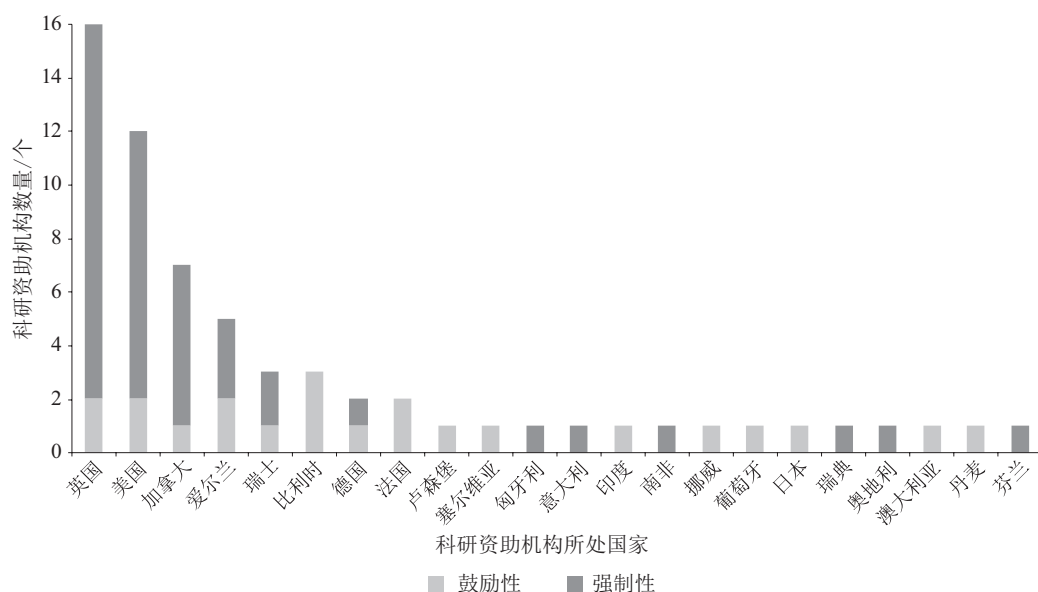


图1 制定科学数据政策的科研机构所属国家分布状况

对登记的42个科研资助机构制定的强制性数据政策进行整理,剔除同一机构修订的政策及机构合并产生的政策变更等情况,共计37个机构制定科学数据政策。然后,依次进入37个国际科研资助机构官网获取政策文本,并结合英国数据监管中心(Digital Curation Centre, DCC)总结的11个资助机构数据政策要素对收集到的37个政策完整性进行评估^[14],最终选取12个较为完善的国际科研资助机构政策作为研究对象,如表1所示。

3 科研机构在科学数据管理流程中的目标分析

3.1 数据计划

科研机构在数据计划阶段应向资助机构提交符合要求的科学数据管理与共享计划,并严格按照计划内容进行数据管理。本研究调研的12个科研资助机构均要求科研人员在项目申请时提交科学数据管理计划,将科学数据管理计划作为项目申请书的一部分。虽然每个机构的计划主题内容存在差异,但基本核心内容

均包括：①数据共享权限与方法；②使用的元数据标准、格式；③数据保存方法、时段，以及所需的设备或基础设施；④预期得到的数据类型、格式；⑤如何保证数据长期开放存取；⑥数据管理的角色和职责。

NSF、NASA明确提出要以审核项目进展报告的形式监督科学数据管理计划的执行。ESRC提出受资助者所属机构，即科研机构应承担其监督受资助者执行计划的责任。

表1 较完善的国际科研资助机构科学数据政策概况

资助机构（英文简称）	国家	政策名称	政策发布或更新时间
美国国家科学基金会（NSF）	美国	《NSF数据共享政策》	2013年1月更新
美国国立卫生研究院（NIH）	美国	《NIH数据共享政策与实施指南》	2003年3月更新
美国宇航局（NASA）	美国	《NASA资助研究成果开放获取政策》	2014年11月更新
艺术和人文研究理事会（AHRC）	英国	《研究资助指南》	2016年4月更新
生物技术与生物科学研究理事会（BBSRC）	英国	《BBSRC数据共享政策》	2007年4月发布， 2016年3月更新
英国癌症研究中心（CRUK）	英国	《数据共享指南》	2009年1月发布
英国国家发展部（DFID）	英国	《DFID研究公开和加强获取政策》	2013年3月发布
英国经济与社会研究理事会（ESRC）	英国	《研究数据政策》	2000年发布， 2015年3月更新
英国医学研究历史会（MRC）	英国	《数据共享政策》	2006年1月发布， 2011年9月更新
自然环境研究理事会（NERC）	英国	《数据政策》	2011年1月发布
英国科学和技术设施理事会（STFC）	英国	《科学数据政策》	2011年9月发布
维康信托基金会（WT）	英国	《数据共享政策》	2010年8月更新

3.2 数据收集

科研机构在数据收集阶段应对项目产生数据及其辅助信息进行采集并在规定时间内向资助机构统一汇交，还应对数据质量承担责任，保证数据真实、完整。本研究调研的12个科研资助机构均要求在规定时间内（项目结束后2个月到2年不等）提交机构资助项目产生的科学数据。其中，NIH、NASA等10个机构要求提交最终数据，即被记录的事实对象，不包括实验室笔记本、部分数据集、初步分析、科学论文草稿、未来研究计划、同行评审报告，以及与同事交流诸如凝胶或实验室标本的物理对象等过程性数据。NSF、NIH、NASA等9个机构明确要求数据背景信息或描述信息应以元数据或数据文档的格式与科学数据集一起提交。其中，NSF要求提交科学数据的派生数据产品和软件，NASA要求提交读取和使用数据所需的软件描述，WT建议提交便于重用的数据格式。大多科研资助机构都要求科研人员自主向指定或推荐的学科或机构数据库提交数据，其中NERC自主建设了数据中心，要求数据中心定期采集科研人员产生的数据。NSF、NASA与ESRC要求

受资助者所属机构应对数据进行审核，以保障数据类型格式规范并防止敏感信息泄露。

3.3 数据加工

科研机构在数据加工阶段应对项目数据进行处理加工，生成完善的元数据以保证后期共享利用。本研究调研的机构中NIH、AHRC、BBSRC、CRUK、DFID、ESRC、MRC、STFC 8个机构均要求生成足够的元数据，一方面便于充分理解科学数据；另一方面规范描述语言，便于整理排序。其中，STFC明确要求元数据应包含对数据权限的细节和存取条件等内容的描述；CRUK要求元数据应包含收集数据的方法、变量的定义、度量单位、数据格式、文件类型等。

3.4 数据长期保存

科研机构在数据长期保存阶段应将具有长期保存价值的依据资助机构要求保存至推荐位置，并保证数据在保存期限内可读可用。本研究调研的12个机

构均要求将科学数据存储至一个开放的学科或机构存储库，其中部分机构指定推荐存储库。如NASA要求将数据存储至PubSpace，AHRC要求将考古数据存储至ADS，NERC要求将数据存储至本机构自建的数据中心。12个机构均对存储时间进行规定，STFC强调不能

重新测量的数据（受其自然属性的影响）应当得到永久保存。除保存最终研究数据与元数据外，NSF、NASA以及ESRC要求资助者所属机构制定数据质量评估标准，鉴定具有长期保存价值的数据进行长期保存，具体如表2所示。

表2 资助机构对数据长期保存的相关要求

资助机构	保存期限	保存类型	保存位置	保存标准
NSF	5年以上	有价值的最终研究数据	合适的数据中心	制定数据质量评估标准
NIH	3年以上	最终研究数据，元数据	指定存储库	-
NASA	5年以上	项目产生的具有长期保存价值的科学数据	PubSpace	制定数据质量评估标准
AHRC	3年以上	最终研究数据，元数据	开放的学科或机构存储库，考古学数据应提交到ADS数据库	-
BBSRC	10年以上	最终研究数据，元数据	能提供永久标识符的存储库	-
CRUK	5年以上	最终研究数据，元数据	开放的学科或机构存储库	-
DFID	5年以上	最终研究数据，元数据	指定存储库	-
ESRC	5年以上	有长期保存价值的最终研究数据，元数据	开放的学科或机构存储库	制定数据质量评估标准
MRC	10年以上，临床或公共卫生研究记录20年以上	最终研究数据，元数据	开放的学科或机构存储库	-
NERC	5年以上	最终研究数据，元数据	NERC数据中心	-
STFC	3年以上，不能重新测量的数据永久保存	最终研究数据，元数据	开放的学科或机构存储库	-
WT	3年以上	最终研究数据，元数据	能提供永久标识符的数据存储库	-

3.5 数据共享利用

科研机构在数据共享利用阶段应最大限度的满足科学数据开放共享需求，并规范和监管不良的数据利用行为。本研究调研的12个科研资助机构均要求以研究成果发布为参照，在研究成果发布前或发布后6个月内共享科学数据集及其元数据或数据文档，其中NSF要求共享派生数据产品。多数资助机构要求科研人员自主按要求共享科学数据，由于NERC自建数据存储库，因此由NERC数据中心在合适时间公开数据。NSF、NIH、NASA等9个科研资助机构提到涉及个人信息、商业秘密等敏感数据应受到保护，共享前要取得受试者知情同意并进行匿名化处理。12个机构均提到利用者应规范引用和使用数据，承认数据来源，以规范格式引用数据。此外，NIH等8个机构提到科研人员可根据数据的敏感性、数据集的大小和复杂性以及预期的需求量选择共享方法，如人与人之间的直接共享或上传至开放存储库的间接共享，具体如表3所示。

4 科研机构在科学数据管理流程中的职责解析

科研机构作为科学数据管理的责任主体，其职责贯穿科学数据生命周期的始终，既是管理者又是服务者更是监督者。针对科研机构在科学数据管理各环节的目标，进一步解析其各环节基本职责如图2所示。

4.1 辅助计划制定，监督计划执行

在计划阶段，科研机构辅助科研人员制定符合科研资助机构要求的科学数据管理计划并监督计划的执行。依据不同科研资助机构数据政策的规定，科学数据管理计划有不同的内容要求与格式规范。为提升科研人员制定科学数据管理计划的质量与效率，科研机构应收集并整理不同资助机构要求，设计具有规范性与可操作性的计划模板；聘任专业数据管理人员对科研人员制定计划提供相关培训或个性化咨询服务。科研机构应要求科研人员定期或不定期提交项目进展报告以

表3 资助机构对数据共享与利用的相关要求

资助机构	共享主体	共享客体	共享条件	共享方法	共享期限	数据利用
NSF	科研人员	数据集, 元数据, 数据产品	敏感数据受到保护	-	最好不晚于研究出版物的发布	以规范格式引用数据
NIH	科研人员	数据集, 元数据	保护参与者的隐私并保护机密和专有数据	依据数据的敏感性, 数据集的大小和复杂性以及预期的需求量选择共享方法, 包括直接共享和间接共享	最好不晚于研究出版物的发布	承认数据来源
NASA	科研人员	数据集, 数据文档, 元数据	敏感数据受到保护	-	最好不晚于研究出版物的发布	以规范格式引用数据
AHRC	科研人员	数据集, 元数据	维护机密, 防范不合理的成本, 保护个人隐私, 尊重同意条款, 以及管理安全性或其他风险	-	在出版物出版前共享, 并应采用可引用的形式	使用他人数据应始终符合法律、道德和监管框架, 包括适当的承认
BBSRC	科研人员	数据集, 元数据	涉及商业机密的科学数据共享受限制	选择适合的共享方法, 包括直接共享和间接共享	不晚于研究出版物的发布	以规范格式引用数据
CRUK	科研人员	数据集, 元数据	涉及人类参与者的研究必须保护个人隐私	选择适合的共享方法, 包括直接共享和间接共享	1年以内	以规范格式引用数据
DFID	科研人员	数据集, 元数据	-	选择适合的共享方法, 包括直接共享和间接共享	最好不晚于研究出版物的发布	以规范格式引用数据
ESRC	科研人员	数据集, 元数据	共享数据前要获得被试者的知情同意, 必要时进行匿名化处理	选择适合的共享方法, 包括直接共享和间接共享	最好不晚于研究出版物的发布	以规范格式引用数据
MRC	科研人员	数据集, 元数据	共享数据前要获得被试者的知情同意	选择适合的共享方法, 包括直接共享和间接共享	最好不晚于研究出版物的发布	以标准格式引用数据或在致谢中加以标识
NERC	NERC 数据中心	数据集, 元数据	-	-	最好不晚于研究出版物的发布	以规范格式引用数据
STFC	科研人员	数据集, 元数据	-	选择适合的共享方法, 包括直接共享和间接共享	研究成果在出版物上发表后的6个月内共享数据	以规范格式引用数据
WT	科研人员	数据集, 元数据	敏感数据受到保护	选择适合的共享方法, 包括直接共享和间接共享	研究成果在出版物上发表后的6个月内共享数据	以规范格式引用数据

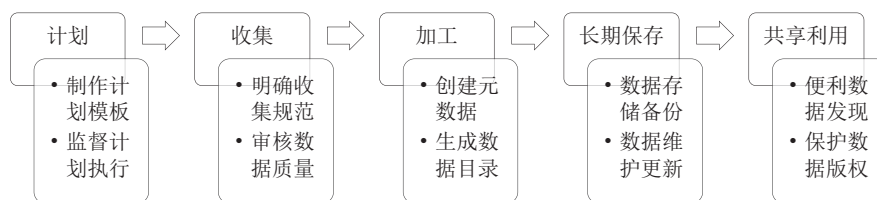


图2 科学数据管理流程中科研机构的职责

审核其是否依据计划进行科学数据管理, 监督并及时纠正数据生产者的数据管理行为。

4.2 明确收集规范, 审核数据质量

在收集阶段, 科研机构应制定本机构内部科学数

据收集规范, 包括明确收集对象范围是项目完成后生成的最终数据或项目进行过程中产生的过程数据; 明确收集内容应包含科学数据集本身以及生成数据的背景信息和解释性信息; 明确收集数据格式以便于长期可读可用与共享传递; 明确收集时间与收集方式, 由数据生产者自主提交或由数据管理者定期采集。科研机构应对收集

的科学数据进行质量审核,包括通过审核其数据产生信息以判断数据的真实性和准确性,审核数据格式是否适合共享和利用,不依赖于特定软件和硬件环境。

4.3 创建元数据记录,生成数据目录

在加工环节,科研机构应生成足够的元数据以支撑之后的数据检索和理解,并对科学数据进行有效组织、整合,生成便于检索、发现的数据目录。为创建元数据记录、生成数据目录,科研机构应在遵守国家与行业元数据标准的基础上,结合自身应用需求,设计元数据方案,据此对科学数据集进行著录与标引,实现数据分类、数据集成与数据关联,并支撑跨平台数据检索与获取。

4.4 存储与备份数据,保障其更新与维护

在长期保存环节,科研机构应根据资助机构要求制定长期保存规范,基于数据价值判断数据存储时限(如5年、10年、20年等),对特殊的不可再生数据进行永久保存。明确长期保存对象应不仅包括科学数据集,还应包括其元数据、解释性数据文档、派生的数据产品以及软件等。科研机构为保障数据长期可读、可用,将科学数据转换至最佳格式并迁移至最佳媒介,进行数据备份,并进行日常的维护与更新。

4.5 保护数据版权,规范引用行为

在共享利用环节,科研机构一方面要最大程度地使数据使用者的数据发现和利用过程简单化、便利化;另一方面要保障数据生产者的知识产权,规范数据使用者的数据利用行为。科研机构应鼓励数据生产者出版或公开发布科学数据及其元数据,但同时保护涉及个人隐私、商业秘密等敏感性数据,在出版前对其进行匿名化处理。科研机构应在分析科学数据使用者行为的基础上,为本机构科学数据管理平台设计用户友好化和多样化的数据发现与数据获取功能,如高级检索功能、数据推荐功能、在线浏览功能、数据分析与可视化功能等。科研机构应通过开设通识课程或系列讲座的方式提升本机构科研人员的数据素养,提升其数据获取能力的同时也提升其规范引用数据的意识,以规范的格式引用数据、合法的方式利用数据。

5 结语

科研资助机构的科学数据政策定义了科研机构在科学数据管理流程的计划、收集、加工、长期保存、共享利用各阶段的基础职责与任务,包括辅助科研人员制定符合资助机构要求的数据管理计划,并监督计划执行;明确数据收集规范,审核收集到的数据质量;为科学数据创建完善的元数据记录,并据此分类整序,生成数据目录;存储与备份科学数据,保障其更新与维护;保护数据版权,规范数据使用者引用行为。科研机构作为科学数据管理的责任主体,明确自身职责与任务对规范和推动本机构乃至国家层面的科学数据管理与共享都具有重要意义。因此,明晰科研机构在科学数据管理流程中的任务是现阶段推动我国科学数据管理与共享发展的突破口。我国科研机构可以以此作为依据和参考,部署和完善本机构的科学数据管理体系,推动我国科研机构科学数据管理与共享的实践进展。

参考文献

- [1] 王丹丹. 新加坡南洋理工大学科研数据管理服务的实施与思考[J]. 大学图书馆学报, 2019, 37(2): 67-74.
- [2] 科学数据管理办法[EB/OL]. [2019-12-14]. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.
- [3] 黄如花, 王斌, 周志峰. 促进我国科学数据共享的对策[J]. 图书馆, 2014(3): 7-13.
- [4] 陈晋. 2008—2018年我国科研数据管理服务研究述评[J]. 图书馆工作与研究, 2019(11): 78-84.
- [5] The Royal Society. Science as an open enterprise[EB/OL]. [2020-01-08]. <https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/>.
- [6] 崔雁. 科学数据开放中科研资助机构研究[J]. 图书馆杂志, 2017, 36(7): 25-36.
- [7] JONES S. Developments in research funder data policy[J]. The International Journal of Digital Curation, 2012, 7(1): 114-125.
- [8] 陈大庆. 英国科研资助机构的数据管理与共享政策调查及启示[J]. 图书情报工作, 2013, 57(8): 5-11.
- [9] 张瑶, 顾立平, 杨云秀, 等. 国外科研资助机构数据政策的调研与分析——以英美研究理事会为例[J]. 图书情报工作, 2015, 59(6): 53-60.
- [10] PINK C. Meeting the data management compliance challenge:

- funder expectations and institutional reality [J]. The International Journal of Digital Curation, 2013 (8): 157-171.
- [11] 黄国彬, 屈亚杰. 英国科研资助机构的科学数据共享政策调研 [J]. 图书馆论坛, 2017, 37 (5): 124-132.
- [12] 邢文明, 汤雅静, 秦顺. 国外教育机构科研数据管理政策大纲解读及启示 [J]. 数字图书馆论坛, 2019 (5): 9-16.
- [13] Sherpa Juliet [EB/OL]. [2019-12-14]. <http://v2.sherpa.ac.uk/juliet/search.html>.
- [14] DCC. Overview of funders' data policies [EB/OL]. [2019-12-14]. <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>.

作者简介

黄源, 女, 1997年生, 硕士研究生, 研究方向: 知识组织与知识服务。

施桐婕, 女, 1997年生, 硕士研究生, 研究方向: 知识组织与知识服务。

李晨英, 女, 1963年生, 研究馆员, 研究方向: 知识组织与知识服务, 通信作者, E-mail: licy@cau.edu.cn。

Analysis of Research Institutions' Responsibilities Based on of Scientific Data Management Workflow

HUANG Yuan SHI XuJie LI ChenYing
(China Agricultural University Library, Beijing 100193, China)

Abstract: In March 2018, the State Council issued the *Measures for the Administration of Scientific Data*, emphasizing that research institutions, as the responsible subjects for scientific data management, should deploy the management and service system of scientific data of their institutions. The scientific data policy of research funding institutions defines the tasks and objectives to be accomplished in the process of managing scientific data, and can effectively promote the scientific data management practice of research institutions. According to the core stages of the scientific data management process, the data policy of the international research funding institutions is analyzed, and the policy requirements are transformed into the responsibilities and tasks of the scientific research institutions, so as to provide reference for the effective deployment of scientific data management in research institutions.

Keywords: Scientific Data; Management Responsibility; Research Institution; Data Policy

(收稿日期: 2019-12-29)