

· 科学论坛 ·

实施公共资助科研项目研究数据 开放共享的政策建议

张晓林*

(中国科学院文献情报中心, 北京 100190; 上海科技大学, 上海 201210)

[摘要] 科研数据开放共享已成为国际科研资助机构的共同要求, 相应的技术条件已经基本具备, 但事实上在很大程度上仍处于“都喜欢但很少做”状态。作者利用新制度经济学原理, 从复杂环境下多元利益相关者的利益调适角度考虑开放共享政策, 提出要分类管理、利益驱动、推拉结合、多措并举、精准施策, 构造统筹协调的科研数据开放共享的良性生态体系。作者针对资助机构、受资助者、资助依托机构、学术团体、数据共享平台、数据使用者等提出具体政策建议。

[关键词] 资助机构; 科研数据; 开放共享; 权益激励与制衡; 政策; 实施指南

研究数据是科学研究进行观察、实验和推理的结果和基础。英国皇家学会 2012 年发布的《科学是一项开放的事业》报告指出^[1], 正是这些实验与观察数据, 支持科技界发现研究失误、支持或否认研究结论、修改完善研究结果、复用研究数据以开展进一步研究。科学的强大的自我纠错能力来源于科学对自身过程与结论的检查与批判的态度和能力, 而这种能力本身依赖于科研数据的开放共享。

科研数据开放共享 (Open sharing 或 Public sharing) 的基础是开放数据 (Open data)。关于什么是开放数据, 科技界已经取得基本一致的认识。前述《科学是一项开放的事业》报告提出, 开放数据应该是可获取 (accessible, 可被发现和获得)、可理解 (intelligent, 可被使用者理解)、可评价 (assessable, 使用者可评价以判断数据可靠性和数据提供者的能力)、可使用 (usable, 可被其他使用者使用)。在互联网时代, 正如《英国开放数据白皮书》^[2] 进一步提出, 开放数据应可以通过互联网获得, 应以计算机可读格式记载, 应对人们使用它或再传播它没有不合理的限制。考虑到数据可能具有不同的开放程度和不同的可关联利用性, Tim Berners-Lee 提出一个五级系统 (Five Stars for linked open data)^[3] 来评价数据的开放共享程度, 根据数据 (1) 是否可在互联网上

获得、(2) 是否是以计算机可读格式提供、(3) 是否使用通用的标准数据格式、(4) 是否用 RDF (Resource Description Framework) 标准进行描述、(5) 是否按照关联开放 RDF 机制进行发布, 分别给予 1 到 5 星等级。最近, 科学数据界越来越认同开放数据的 FAIR 原则^[4], 即数据必须是可发现 (Findable)、可获取 (Accessible)、可互操作 (Interoperable) 和可复用 (Reusable)。

科研数据开放共享本身往往是一系列政策、机制、工具、平台等集合所形成的过程与行为^[5], 需要开放、灵活、透明的机制, 与现行法律体系相一致, 对知识产权予以合理保护, 支持科研管理和科研数据管理的专业性, 支持数据的质量管理和可互操作性, 提高数据管理与共享的安全与效率, 还需要对数据共享进行明确赋责和赋能、保障数据共享的可问责性和可持续性。这些往往要求一系列的甚至体系化的数据管理服务^[6]。

1 科研数据开放共享的发展现状

1.1 科研数据开放共享是科技创新与社会发展的需要

科研数据开放共享^[7] 不仅有效地支持对科研成果的检验和重复验证, 还支持人们根据新的理论或

收稿日期: 2018-12-31; 修回日期: 2019-01-03

* 通信作者, Email: zhangxl@mail.las.ac.cn

方法对数据进行新分析、或将数据与其他数据融汇后进行不同的分析,支持人们按照新的视角、场景或目的重新利用数据。而且,科研数据是重要的知识资产,包含丰富的知识对象,它们的开放共享以及集成,支持对科研大数据的深度挖掘,支持把科研数据与其他类型的社会、经济、市场、环境、文化、健康、安全、管理等方面的感知信息相互关联和融汇,支持智慧化管理和智慧社会建设。科研数据开放共享是建设和支撑这个智慧社会的重要工具。

如果缺乏有效的科研数据管理和共享,科研人员很难询证科研成果和支持重复验证,难以支撑可靠的合作研究,难以证明自己的研究质量和机构的研究管理质量。科研机构不能保护科研数据,不能积累数据资产,不能汇集数据开展新的研究,可能因为管理失灵而造成数据丢失、管理失范造成数据即使保存下来也无法使用。对中国来说还涉及数据安全,中国科学家在国外期刊上发表论文时越来越多地被要求提供支撑数据,但却不得不把这些数据存放到期刊所指定的国外存储库中,而作者所在的中国机构和投资科研的基金机构却无法保存数据。

1.2 科研数据开放共享已经成为普遍的国际趋势

国际范围内,科研数据开放共享已经得到长足发展,逐步成为公共资金资助的科学研究的基行为准则,相关的发展趋势可总结为:

多数国家的公共科研资助机构已制定了数据管理与共享政策,例如美国国立健康研究院的数据共享政策^[8],美国科学基金会的数据管理与共享政策^[9],英国研究与创新理事会的数据共享原则^[10],欧盟地平线 2020 计划的数据共享要求^[11]等等。

越来越多的学术期刊要求研究论文向科技界乃至社会共享论文的支撑数据,例如《科学》(*Science*)针对论文支撑数据^[12],要求在论文发表前这些数据必须存储到得到刊物同意的数据存储库中并获得一个数据集标识,这个标识必须被包含在论文中;论文发表后,所有有助于理解、评价和扩展论文研究结论的数据,要对任何提出合理索取要求的读者提供。目前,《自然》(*Nature*)^[13]、《细胞》(*Cell*)^[14]、Springer Nature^[15]、Elsevier^[16]、Wiley^[17]、Taylor and Francis^[18]等刊物或出版社都已明确提出数据开放共享要求,部分学术期刊提出了自己认可的数据存储库清单^[19]。

不断增多的数据期刊支持发表数据文章,例如 Scientific Data^[20]、GIGA Science^[21]、《中国科学数

据》^[22]等;不断增多的可公共共享的数据存储库已经建立,例如 FigShare^[23]、Dryad^[24]等,按照 Re3Data^[25]统计,目前已有 2 000 余个通用的或专业的公共数据存储库。

同时,越来越多科研教育机构建立起自己的科研数据管理与共享制度,例如剑桥大学^[26]、牛津大学^[27]、加州大学伯克利分校^[28]、哈佛大学(包括数据管理服务^[29]和数据安全规范^[30])、麻省理工学院^[31]、斯坦福大学^[32]、耶鲁大学^[33]等。

欧盟还建立了专门的开放科学监测平台(EU Open Science Monitor)^[34],对各成员国执行论文开放获取和数据开放共享政策的情况定期进行监测报告。

1.3 科研数据开放共享的技术条件已经基本具备

通过科技界的共同努力,支持科研数据开放共享的许多业务规范、技术标准、流程指南、最佳实践等“技术条件”已经基本具备。

例如,需要在科研项目开始前就对项目研究数据的管理进行规范细致的规划,为此已经有 DMPTool^[35]和 DCC 的 DMP^[36]两个平台提供了大量可借鉴的通用的或专门的数据管理计划模板、甚至在线辅助规划服务。

又如前述,科技数据界提出了开放数据的 FAIR 原则,而且对如何在技术和流程上保障数据 FAIR 化提出详细要求,例如 W3C 提出的 Data on the web best practices 标准^[37]。

要做到有效的研究数据管理,人们提出把数据管理的机制嵌入到科研过程中,形成基于项目生命周期的科研数据管理机制^[38]。而且,新型的数字化实验室记录本系统(Electronic Lab Notebook)已能较好地在研究过程中管理数据的产生、记录、处理、传递、发布等过程^[39]。

为了有效共享研究数据,人们已建立了数据引用(Data Citation)规范^[40],不断优化各类科研数据描述标准^[41],提出了网络调用数据的公共协议(OData Protocols)^[42],推出了数据开放共享授权协议(Open Data Common Licenses)^[43],提出了公共可共享数据存储库的建议名单^[44]和评价这类存储库的参照标准^[45]。有关刊物还提出在发表论文时作者需要做出的数据可获得性声明规范^[46]。已经有出版社向作者和研究社群提供研究数据管理服务和开放数据咨询服务(例如 Springer Nature 的 Research Data Support^[47]和 Open Data Help Services^[48])。

2 科研数据开放共享需要综合协调的利益激励与制约机制

2.1 科研数据开放共享仍然面临严峻的挑战

尽管科研数据开放共享的意义已得到普遍认同、相应技术条件也基本具备,但事实上科研数据开放共享在很大程度上仍处于“都喜欢但很少做”的状态(everybody loves data sharing, but nobody does it)^[49]。据2018年Springer Nature对7700多名研究人员的调查^[50]、Wiley在2016年针对112个国家4600多名研究人员的调查^[51]和Digital Science对开放数据现状的调查^[52],多数研究人员认为数据共享还远远没有达到应有的程度,他们自己还仅仅是在小范围进行数据共享、往往仅仅因为期刊的要求而公布数据、往往缺乏进行数据监护(data curation)和数据共享的能力、激励和可持续支持。

人们认识到,科研数据本身以及科研数据开放共享涉及复杂的因素,不是简单的非黑即白的情景,需要As open as possible, as closed as necessary。Sarah Jone分析了科研数据管理(RDM)、开放数据(Open Data)和FAIR原则的关系^[53],指出三者并非重合。科研项目需要对所有数据进行有效管理,但这些数据中可能只有部分数据(例如多层次结果数据而非原始裸数据)需要按照FAIR原则组织,而且只有部分裸数据和FAIR化的结果数据会作为开放数据提供共享。

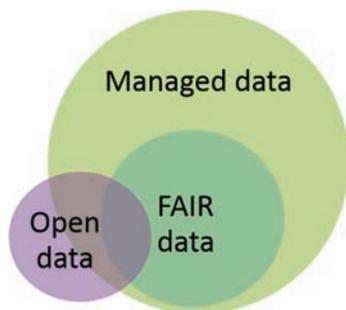


图1 RDM/FAIR/OD关系

《科学是一项开放的事业》中提到,保密、隐私保护、商业竞争等都是某些数据不开放共享的合理理由。英国研究与创新理事会的《研究数据管理最佳实践指南》^[54]进一步分析,要考虑数据共享的复杂制约因素,例如关于安全、隐私、商业竞争、学术研究竞争、数据采集与处理方法创造性、以及是否是自己产生和拥有的数据、或者数据的全部或部分来自别人或合作者;要考虑数据共享的成本及其影响,例如在数据的整理、描述、保存、更新、管理、服务等方面

的直接成本,在精细规划和实施数据管理与监护过程中的间接成本,以及按科研生命周期进行数据采集、处理、转换、保存、传递等时需要的方法、工具、流程、人员等的间接成本;要关注如何激励和保护数据共享的意愿,例如对数据拥有者、管理者和共享服务者的权益保护、对数据共享的激励措施、将数据共享纳入科研贡献评价和科研绩效评价、以及对数据的共享使用可能出现的各种情况的免责保护;还要考虑如何提供数据共享的支持条件,包括知识、技能、技术方法、平台与工具、经费、人员等条件。

Houtkoop等人对心理学领域研究人员的共享认知进行了调查^[55],发现相当多的人觉得在自己领域里开放共享还没有成为普遍行动(68%),超过一半的人认为按共享要求来准备数据非常费事(55%)、不知道如何在线进行共享(54%),近三分之一的人感觉缺乏对共享的认可和激励(27%),还有少数人(12%)认为没有合适的数据存储库来发布自己的数据。进一步分析发现,即使对希望共享自己数据的人,提供可靠和方便的共享基础设施是实现共享的基本条件;对那些愿意共享自己数据的人,将共享所需工作嵌入科研流程是很重要的促进条件;更多的人需要得到所在领域关于数据共享最佳实践指南的详细指导;处于繁忙科研竞争中的人需要对数据共享的激励措施;对少数希望固守现有状态的,则需要明确的政策去要求和监测开放共享行动。

2.2 科研数据开放共享需要复杂的利益调适机制

公共政策实质上是对复杂条件下相互博弈的多元利益相关者的利益调适机制。我们已知,科研数据开放共享涉及社会、科技界、资助者、科研机构、科研人员、数据人员、出版者、基础设施提供者等。开放共享正是对这些利益相关者的复杂利益进行积极、精准、平衡的调适的结果,简单地忽略或维护、剥夺或加强某些利益相关者的利益的做法都很容易导致共享的弱化甚至失败。

这里,新制度经济学对我们应有启发^[56]。它认为:人的行为是有限理性的;人都有为自己谋最大利益的机会主义行为倾向。这种情况下,人们怎样才能做好事呢?人们做好事的行为认知基础是什么?初步总结,包括:从道义上觉得应该做,从心理上愿意做,从支撑条件上能够做,从激励、评价和奖惩机制上知道做了会有积极作用、不做会有消极影响,从操作上应该方便自然不费劲。

为了达到上述效果,科研数据开放共享的实施

政策需要分类管理、利益驱动、推拉结合、多措并举、精准施策;政策措施需要可操作、可测量、可验证、可比较、可问责;实施行动上需要能做就做、能做先做、互相铺垫、逐步合围,积极主动又统筹协调地建立起有利于和促进科研数据开放共享的生态机制。

2.3 科研数据开放共享需要分类管理

“科研数据”是个复杂体,“开放共享”也是个复杂现象,不可能也不应该有个“一刀切”的政策及其实施机制,需要根据共享的数据、对象、程度、使用许可、开始共享时间等的不同进行分类管理。

(1) 数据,可分为:论文结果的直接支撑数据;论文的其他支持数据;与项目相关的其他研究数据;与项目相关的工具、方法、系统等的数据。不同类别的数据在安全保护、竞争需要、共享成本、共享需要等上存在很大区别。

(2) 共享对象,可分为:本机构或合作机构科研人员;本学科科研人员;其他科研人员;有能力和需求的企业机构或公众;社会公众。很明显,不同的对象往往在不同时间对不同类别的数据有不同的使用需求。

(3) 数据的共享程度,以一个覆盖多个数据指标的长期积累的数据库为例,可分为:共享份量(少部分、一定比例、多部分、全部?)、共享频次(一次、多次但可预期、常常且不可预期、随时?)、共享用途(结果验证、科研复用、政府管理、商业应用等)。可以理解,对于不同程度的共享往往意味着不同的政策、许可、手续、成本、方法等。

(4) 共享的管理机制,可以是:完全开放;开放但进行监测;开放但需要登记;开放但需要申请;开放但要缴费(回收成本或者有所盈利?)等。

(5) 共享的启动时间,可以在项目期间就可共享、项目结束后才能共享、项目结束后一定时间才能共享等等。研究人员往往有理由根据不同的数据、共享对象、共享程度等设置不同的共享启动时间和共享管理机制。

2.4 科研数据开放共享需要保护合法要求和利益

只有在人们的合法要求和合理利益得到保护时才会做出社会所期待的行为,尤其是在这种行为可能意味着人们做出额外的努力或需要让渡部分利益时。因此,在科研数据开放共享中,需要充分认可对共享的限制性处理。例如:

(1) 认可对共享的某些正常限制,包括法律、道德(隐私、私密性、知情同意)等的限制,以及出于正

当商业需要的共享限制。

(2) 认可某些情况下需要延后数据共享,例如应支持项目团队在有限时间内把数据仅限于内部使用,支持项目团队撰写与发表论文;如果项目团队花很大精力用于数据采集,一般应允许更长的内部使用;如果准备数据需要较长时间,因此适当延长数据开放也是合理的;而且,往往为了不影响还在进行中的科研、或者需要全谱段数据才能可靠分析,则应避免过早共享数据。

(3) 认可需要保护数据采集、管理和共享中的知识产权;大规模、长时间的数据采集和处理体现了劳动贡献,或者在数据采集或处理或服务中创造了新方法新技术,可允许留出有限的数据暂不共享时间,支持撰写数据文章、或方法文章、或系统平台文章等,并要求共享后的数据使用者引用这些文章,使相应贡献得到学术上的承认。

(4) 认可可靠的数据管理与高效的数据共享需要成本,设计合适的机制来回收或补偿相应的成本,支持数据管理与共享的可持续进行。

(5) 承认对不同用户、对不同数据的不同用途需要不同的管理机制,需要针对性地建立数据共享的条件与要求。例如,欧盟开放科学政策平台专家组就开放共享的激励与保护措施提出的建议可供我们参考^[57]。

从操作层面上讲,可以有:(1) 让数据集可在学术文献中被引用;(2) 赋予数据集一个唯一数字标识符(DOI);(3) 用规范元数据描述这个数据集、并把它登记到一个可公共发现和获取的数据仓储库中;(4) 按照 FAIR 原则处理数据集使得其可被发现和利用;(5) 针对数据集提供数据共享的条件与许可;(6) 对那些首家利用数据、大规模使用数据、商业化使用数据、将数据集作为某种应用的常态化基础设施(DaaS)等,设定有针对性的许可条件和管理机制;(7) 尽快针对数据收集、管理与共享服务撰写自己的学术论文,例如数据论文、方法论文、工具或系统论文等;(8) 定义针对不同程度的数据共享服务而需要的不同权益,例如要求使用者致谢、要求使用者引用自己与被使用数据密切相关的论文、在提供了复杂专门个性化服务时作为研究成果的合作作者等;(9) 获得对日常数据管理与共享服务的可持续性支持,包括相应的技能、知识、人员和资助;(10) 建立合适的监测与评价制度,使得数据共享可测量可检验可问责,也使得数据共享工作在科研评价和工作绩效评价中得到认可。

2.5 开放共享政策需要在利益激励与制衡上推拉结合

在复杂条件下对多元利益相关者的利益进行调适,需要在利益的激励与制衡上采取推拉结合的机制^[58]。毕竟,对于复杂的利益关系,权利和利益等的边界常存在模糊之处,权利之间存在博弈,总是存在一些“正面支持因素”,即能够为某些权利人带来正面利益、从而支持权利行使的因素,也往往存在一些“负面反对因素”,即可能对某些权利人带来负面影响、从而遭致对权利行使进行反对的因素。因此,在政策设计时,应注意建立能够强化正面支持因素和消减负面反对因素的措施。例如,在科研数据开放共享环境下,强化正面支持因素的措施可能包括:(1) 激励类:即“做了有好处”的激励措施,例如要求使用者规范引用数据集、鼓励数据提供者撰写数据文章、提供数据集利用统计等;(2) 能力类:即“能够做、方便做”的措施,帮助方便流畅进行数据共享,例如提供数据管理计划模板和数据共享许可模板,提供技能培训,提供可持续经费资助;(3) 服务类:即“从条件上支持做、支持做好”的措施,例如建立可公共共享的数据仓储库,提供数据集发现引擎,提供数字化实验室记录本工具等。消减负面反对因素的措施可能包括:(1) 保护权利类:即“做了没坏处”的保护措施,允许在合理原因下不共享或暂不共享数据,提供法律上有效的共享许可机制;(2) 尽责管理类:即“承担要做的责任”的措施,在科研资助时规定受资助者的义务,对资助依托单位规定进行督促、管理和检查的责任等;(3) 评价类:即“不做有害处”的措施,对数据管理和数据共享进行监测和公告,要求没有尽责实施数据共享的人或机构在评价、资助上承担负面效果。

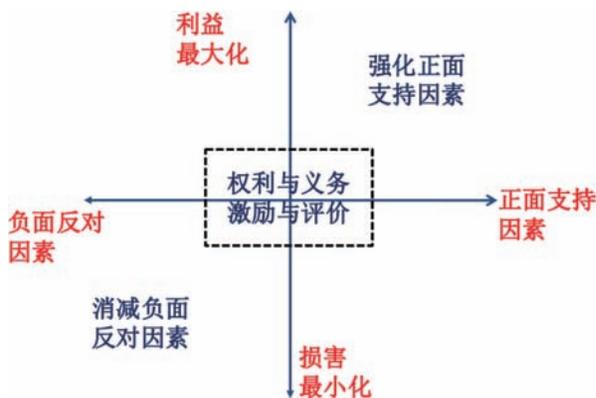


图 2 利益调适的推拉结合机制

2.6 开放共享政策需要多措并举和精准施策

科研数据开放共享政策不仅需要推拉结合,还要考虑通过政策宣示、行为要求、标准规范、流程指南、支撑设施、技能培训、服务咨询等多种措施来创造数据开放共享的良好生态。

而且,复杂系统下的公共政策需要精准施策,针对不同主体(课题组长、科研机构、科技领域、设施提供者、成果出版者、数据使用者等)制定不同政策,针对在规范、流程、技能、设施、服务、评价和管理等不同需求制定不同政策,针对各类主体“最敏感需求”和“痛点”制定相应政策,例如对受资助人和资助依托机构,数据共享的合规性应作为资助的必要条件和后续资助的检查条件,而对数据工作者,激励与评价就变得很重要。

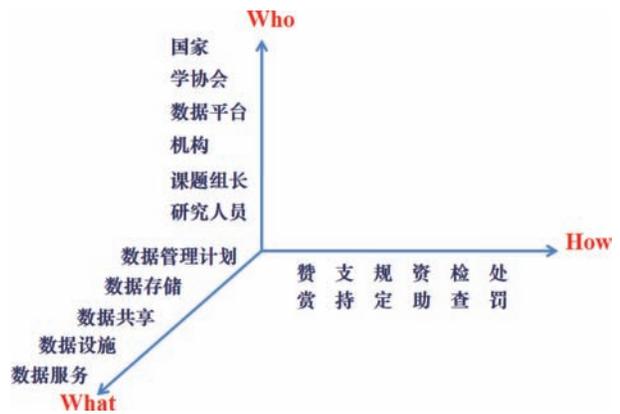


图 3 数据开放共享政策多措并举和精准施策框架

进一步的,要将政策实施的责任压力以合适的方式传递到合适的主体。例如,科研资助机构往往面对成千上万的受资助者、且对其缺乏直接行政管理能力,因此可作为资助条件之一对资助依托机构提出管理和检查数据共享政策实施情况,把机构利益与它所管理的受资助人行为绑定。而且,还有采取多种“推进”措施,尤其是可测量可检验可问责的公开机制,包括报告、监测、审计、公布、整改、处罚等机制。只有在做了有好处、不做有害处时才会去做,只有在做了后得积极效果和不做时得到消极效果才能确证去做的智慧,从而形成良性循环。

3 公共科研资助机构实施科研数据开放共享的政策建议

科研资助机构应该建立和发布正式的科研数据开放共享政策:明确支持科研数据开放共享;明确“科研数据默认为开放共享”,如果不开放共享则需要申明和说明原因;明确要求利益相关者履行各自

主体责任、实施数据开放共享;明确要设计合适的政策实施机制来激励和约束受资助者、资助依托机构、数据平台、出版社、使用者等。

科研资助机构应该资助各个科研机构、专业领域、第三方服务机构等在公开、公平、择优竞争、长期监督检验的条件下建立良好的开放共享生态环境所需的工具与设施,包括但不限于数据管理模板、数据描述与组织规范、数据共享使用许可模板、数据仓储库和数据发现引擎、科研人员培训、数据人员培训等。

科研资助机构应该明确建立对科研数据开放共享的合规监测、对标比较、奖优罚劣机制等,有些可以通过第三方服务进行。科研资助机构应该参与国内各类资助机构的科研数据开放共享的政策与实施的协调努力,积极参加国际资助机构的科研数据开放共享的政策研究、工具与设施建设、监测与评价等努力。

3.1 对受资助者的政策要求

科研资助机构应该明确规定研究项目的受资助者是科研数据开放共享的主体责任人,要求:(1)项目负责人对项目数据的管理与共享负责任;(2)项目申请时必须提交规范的数据管理计划;(3)项目进行过程中采用基于项目生命周期的数据监护系统;(4)允许项目经费支持数据监护和共享工作;(5)项目的阶段性检查和结题验收时必须按照规范形式报告和检查数据共享情况。

科研资助机构应该进一步要求:(1)项目研究数据要提交到可信赖的机构或专业领域的数据仓储库中;(2)项目提交的研究数据应符合 FAIR 要求;(3)项目要明确说明研究数据的开放共享条件与要求,否则将默认执行资助机构规定的开放共享条件与要求。

科研资助机构还可以要求:(1)项目发表论文要同时发布规范的数据可获得性声明;(2)项目发表论文时所提交的支撑数据必须存放一份在中国境内合规的数据仓储库中;(3)支持项目发表数据论文、方法论文和平台论文,所需经费可由项目经费支付;(4)支持项目建立公开的规范的数据共享服务根据贡献度的权益要求声明,允许项目按照成本回收原则收取一定服务费用。

3.2 对资助机构的政策要求

科研资助机构应该明确规定研究项目的依托管理机构也是科研数据开放共享的主体责任人、应对本机构受资助者的数据共享行为负责,要求依托机

构:(1)建立本机构的科研数据管理与共享制度;(2)建立对数据开放共享中选择不公开或暂不公开的申请的审核机制;(3)定期检查和向资助机构报告受资助项目数据开放共享情况;(4)建立科研数据开放共享的评价与激励机制;(5)建立或委托建立可信赖的数据仓储系统。

科研资助机构应该进一步要求依托机构:建立或委托建立科研数据开放共享的知识体系,包括数据管理计划模板、数据可获得性声明模板、数据共享使用许可协议模板、数据共享服务根据贡献度的权益要求模板、数据开放共享中的安全管理规则、数据开放共享中的隐私保护规范、数据共享中的免责声明模板等。

科研资助机构还可以要求依托机构:(1)提供基于项目生命周期的数据监护系统支持;(2)向受资助者提供数据监护与共享的培训;(3)建立支持数据管理与共享的服务团队。

3.3 对数据平台的政策要求

科研资助机构应该鼓励和支持建立合规的数据平台来保存科研数据和提供共享服务。科研资助机构应该对承担这类任务的数据平台提出明确要求:(1)提供可靠的互联网接入,保障健壮的对所存储数据的公共检索、发现和调用服务;(2)采用相关领域认可的数据质量评价机制;(3)支持相关领域通用的数据格式和元数据;(4)支持数据 FAIR 化,提供采用规范标准建立的 OpenAPIs 来方便支持数据调用;(5)建立严格的安全保护机制和权利保护机制,建立支持精细许可与授权的能力;(6)承诺不排他性地利用所存储数据进行自己的商业化或学术化开发;(7)平台本身的政策、流程、规范等提供开放获取,保证可监测可验证;(8)定期向科研资助机构报告和向社会公布自己保存的数据,报告和公布数据共享情况;(9)拥有技术、法律上进行数据管理和共享服务的合适人员;(10)以免费的或非赢利的机制,向所服务对象提供数据管理与数据共享的服务。

3.4 对专业学术团体的政策要求

科研资助机构应该鼓励和支持专业学协会或其他类型的专业学术团体积极参加和促进科研数据的开放共享,包括:(1)建立本领域的科研数据开放贡献的共同原则、实施指南和最佳实践案例;(2)建立和推行符合本领域特点的数据管理计划、数据可获得性声明、数据共享使用许可协议、数据共享服务根据贡献度的权益要求、数据共享安全管理、数据共享隐私保护、数据共享免责声明等的规范或模板及实

施细则;(3) 建立和推行符合本领域特点的数据格式和元数据规范;(4) 建立或委托建立领域的数据平台,并对其合规运行进行监督;(5) 建立自己的开放数据培训与咨询服务,对本领域人员的数据共享进行指导和支持。

3.5 对数据使用者的政策要求

科研资助机构应该鼓励和支持使用者对科研数据的发现、遴选、调用和使用,同时也需要提出数据共享使用的行为规范,包括:(1) 严格遵守共享使用许可的要求,维护数据提供者和服务提供者的合法权益;(2) 主动支持对数据权、隐私、安全等的保护,发现开放数据中存在这方面问题时及时向数据提供者报告;(3) 主动支持对数据及其采集处理方法的检验,发现开放数据中存在这方面问题时及时向数据提供者报告,必要时向其所在机构或资助机构报告;(4) 诚信尽责地支持数据平台和共享服务系统的安全高效运行,不采用可能危害信息系统安全或服务效率的操作行为;(5) 尽可能开放共享自己利用开放数据产生的研究结果和数据,包括但不限于将其存储到可开放共享的数据仓储库中;(6) 支持数据共享服务的可持续运行,承认数据提供者或数据平台在提供深度服务时的合理权益要求,支付合理的服务费用;(7) 支持对数据开放共享的评价和监督,积极提供反馈信息和能促进更好共享的建议与意见;(8) 积极参与数据开放共享的合作社群以及向其他数据提供者和使用者提供帮助。

3.6 对资助机构自身的政策要求

科研资助机构应该致力于营造科研数据开放共享的积极态势和可持续发展生态环境。因此,科研资助机构应该:(1) 将相应政策纳入对申请资助者的申请指南和资助合同中;(2) 项目中期检查和结题验收时要求受资助者进行报告、要求检查者验收者进行核查;(3) 要求项目依托机构定期报告各自实施科研数据开放共享的可测量可检验可问责的结果;(4) 定期公布各项目依托机构实施科研数据开放共享的情况,并对在实施中存在严重问题的机构进行处罚,包括暂停该机构人员申请资助的权利或削减该机构人员已获得资助的经费。

同时,科研资助机构还应该:(1) 资助各主要学科领域研究建立科研数据开放共享实施指南和最佳实践案例;(2) 资助建立规范的数据管理计划、数据可获得性声明、数据共享使用许可协议、数据共享服务贡献度权益要求、数据共享安全管理、数据共享隐私保护、数据共享免责声明等的规范或模板及实施

细则;(3) 资助建立或优化符合 FAIR 原则的数据格式和元数据规范;(4) 资助建立各主要学科领域的数据库;(5) 资助开展针对科研数据开放共享的多层次培训,资助建立相应的咨询服务机制。

2018年3月,国务院办公厅发布《科学数据管理办法》,明确提出要“科学数据管理遵循分级管理、安全可控、充分利用的原则,明确责任主体,加强能力建设,促进开放共享”,明确要求国家科学技术行政部门、各类主管部门和各类法人单位建立健全科学数据相关管理制度、科学数据平台或管理系统、积极开展科学数据共享服务。希望本文提出的相关分析与具体措施能为将科学数据管理与开放共享尽快落到实处提供参考,真正通过可操作、可测量、可检验、可比较、可问责的机制来把好事做快、做到、做好。

参 考 文 献

- [1] Royal Society. Science as an open enterprise. 2012. <https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/>.
- [2] Minister of the Cabinet. UK Open Data White Book, 2012. https://data.gov.uk/sites/default/files/Open_data_White_Paper.pdf.
- [3] Berners-Lee, Tim. Five Stars for linked open data. <http://www.w3.org/DesignIssues/LinkedData.html>.
- [4] Wilkinson MD, Dumontier M, Aalbersberg I J J, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 15 March 2016.
- [5] UIRI. Concordat on Open Research Data. 2016. <https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf/>.
- [6] Bryant R, Lavoie B, Malpas C. The Realities of Research Data Management. Part One: A Tour of the Research Data Management (RDM) Service Space. OCLC Research, 2017. <https://www.oclc.org/content/dam/research/publications/2017/oclcresearch-research-data-management-service-space-tour-2017.pdf>.
- [7] OECD. Principles and Guidelines for Access to Research Data from Public Funding <http://www.oecd.org/sti/inno/38500813.pdf>.
- [8] NIH. Data Sharing Policy and Implementation Guidance. 2003 https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm.
- [9] NSF. Dissemination and Sharing of Research Results. 2010 <https://www.nsf.gov/bfa/dias/policy/dmp.jsp>.

- [10] UKRI. Common principles on data policy. <https://www.ukri.org/funding/information-for-award-holders/data-policy/common-principles-on-data-policy/>.
- [11] EU. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, 2017. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.
- [12] Science Journal; Editorial Policy. <http://www.sciencemag.org/authors/science-journals-editorial-policies>.
- [13] Nature. Availability of data, materials, code and protocols. <http://www.nature.com/authors/policies/availability.html>.
- [14] Cell. Information for authors. <https://www.cell.com/cell/authors>.
- [15] Springer Nature. Research Data Policy Types. <https://www.springernature.com/gp/authors/research-data-policy/data-policy-types/12327096>.
- [16] Elsevier. Research Data. <https://www.elsevier.com/about/policies/research-data>.
- [17] Wiley. Data Sharing Policy. <https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html>.
- [18] Taylor and Francis. Understanding our data sharing policies. <https://authorservices.taylorandfrancis.com/understanding-our-data-sharing-policies/>.
- [19] See the list of databases included in the Data Deposition section in Science Journals; Editorial Policies. <http://www.sciencemag.org/authors/science-journals-editorial-policies>.
- [20] Scientific Data. <https://www.nature.com/sdata/>.
- [21] GiGA Science. <https://academic.oup.com/gigascience>.
- [22] 中国科学数据. <http://csdata.org/>.
- [23] Figshare. <https://figshare.com/>.
- [24] Dryad. <http://www.datadryad.org/>.
- [25] Re3Data. <https://www.re3data.org/>.
- [26] University of Cambridge. Research Data Management Policy Framework. <https://www.data.cam.ac.uk/university-policy>.
- [27] Oxford University. Policy on the Management of Data Supporting Research Outputs. <http://researchdata.ox.ac.uk/university-of-oxford-policy-on-the-management-of-data-supporting-research-outputs/>.
- [28] University of California, Berkeley. Berkeley Research Data Management. <http://researchdata.berkeley.edu/>.
- [29] Harvard University Library. Research Data Management Program. <https://hlrdm.library.harvard.edu/>.
- [30] Vice Provost Office. Harvard Research Data Security Policy (HRDSP). <https://vpr.harvard.edu/pages/harvard-research-data-security-policy>.
- [31] MIT Data Management. <https://libraries.mit.edu/data-management/>.
- [32] Stanford University. Data Management Services. <https://library.stanford.edu/research/data-management-services>.
- [33] Yale University. Yale Open Data Access (YODA) Project. <http://yoda.yale.edu/welcome-yoda-project>.
- [34] EU Open Science Monitor. Facts and Figures for open research data. https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/facts-and-figures-open-research-data_en.
- [35] DMPTool. <https://dmptool.org/>.
- [36] Data Curation Center Data Management Plans. <http://www.dcc.ac.uk/resources/data-management-plans>.
- [37] Data on the Web Best Practices. W3C Recommendation 31 January 2017. <https://www.w3.org/TR/2017/REC-dwbp-20170131/>.
- [38] 刘峰. 面向科研项目全生命周期的科学数据监护框架研究. 中国科学院大学博士论文, 2016.
- [39] The Electronic Lab Notebook in 2018: A comprehensive guide. <https://www.labfolder.com/electronic-lab-notebook-eln-research-guide/>.
- [40] Data Citation Synthesis Group. Joint Declaration of Data Citation Principles, 2014. <https://www.force11.org/group/joint-declaration-data-citation-principles-final>.
- [41] DCC. List of Metadata Standards. <http://www.dcc.ac.uk/resources/metadata-standards/list>.
- [42] OData Version 4. 01. Part 1: Protocol. Committee Specification 01. 30 January 2018. <http://docs.oasis-open.org/odata/odata/v4.01/odata-v4.01-part1-protocol.html>.
- [43] Open Data Commons. <https://opendatacommons.org/index.html>.
- [44] NIH. Data Sharing Repositories. https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html.
- [45] PLOS. Criteria for Recommended Data Repositories. <https://blogs.plos.org/everyone/2018/03/01/criteria-for-recommended-data-repositories/>.
- [46] Taylor and Francis. Data availability statement templates. <https://authorservices.taylorandfrancis.com/data-availability-statement-templates/>.
- [47] Springer Nature. Research Data Support. <https://www.springernature.com/gp/authors/research-data-policy>.

- [48] Springer Nature. Data Helpdesk. <https://www.springernature.com/gp/authors/research-data-policy/helpdesk/12327114>.
- [49] Felix Schönbrodt. Changing incentive structures to foster the actual sharing rate of open data. Open Science Conference 2018. 3, Berlin, Germany.
- [50] Springer Nature. Practical Challenges for Researchers in Data Sharing. 2018. 03 <https://www.springernature.com/gp/open-research/open-data/practical-challenges-white-paper>.
- [51] Wiley. Global Data Sharing Trend in Open Science Trends You Need to Know About. <https://hub.wiley.com/community/exchanges/discover/blog/2017/04/19/open-science-trends-you-need-to-know-about>.
- [52] Digital Science. The State of Open Data 2018. <https://www.digital-science.com/resources/portfolio-reports/state-open-data-2018/>.
- [53] Sarah Jone. Open data, FAIR data and RDM: the ugly duckling. Open Science Conference, 2018. 3, Berlin, Germany.
- [54] UKRI. Guidance on best practice in the management of research data. <https://www.ukri.org/files/legacy/documents/rcukcommonprinciplesondatapolicy-pdf/>.
- [55] B. L. Houtkoop, C Chambers, M Macleod, et al. Data Sharing in Psychology: A Survey on Barriers and Preconditions. *Advances in Methods and Practices in Psychological Sciences*. 2018, 1(1): 70—85.
- [56] 柯武刚, 史漫飞著, 韩朝华译. 制度经济学—社会秩序与公共政策. 北京: 商务印书馆, 2000, 11.
- [57] EU OSPP. Open Science Policy Platform Recommendations, 2018. 04. https://ec.europa.eu/research/openscience/pdf/integrated_advice_opspp_recommendations.pdf.
- [58] 张晓林, 张冬荣, 李麟等. 机构知识库内容保存与传播权利管理. *中国图书馆学报*, 2012, 7: 46—54.

Policy Recommendations for Public Sharing of Research Data from Publicly Funded Research Projects

Zhang Xiaolin

(National Science Library, Chinese Academy of Sciences, Beijing, China; ShanghaiTech University, Shanghai, China)

Abstract Public sharing of research data (data sharing) is now the common drive from research funders world-wide, and the technical supports needed are basically in place. However, data sharing is still to a large degree that most like it but few does it. The author utilizes the principles of the New Institutional Economics, economics, explores the framework of the data sharing policy as a rights adjustment mechanism for multiple stakeholders under a complex gaming system. The author develops detailed requirements for the policy to be differentiating according to varying needs, attending to interests of those involved, combined with encouraging and discouraging pushes, implemented with multiple measures affecting various stakeholders with focused targeting effects, for building a coordinated, positive, and sustainable data sharing ecosystem. The author then outlines the policy recommendations for funders, funded researchers, funded institutions, learned societies, data sharing centers, data users.

Key words research funders; research data; public sharing; incentives and refrains; policy; implementation guidelines