

科学数据开放共享已经成为共识, 欧盟《OECD关于公共资助的科学数据获取的指导方针和原则》、英国皇家学会《科学是开放事业报告》、美国白宫科技政策办公室开放政府政策、欧盟 Horizon 2020 计划都将科学数据作为一种重要的科研产出, 提倡将科学数据开放, 供科研人员获取、挖掘、利用。我国在2015年8月发布的《国务院关于印发促进大数据发展行动纲要的通知》指出, 要发展科学大数据, 积极推动由国家公共财政支持的公益性科研活动获取和产生的科学数据逐步开放共享, 构建科学大数据国家重大基础设施, 实现对国家重要科技数据的权威汇集、长期保存、集成管理和全面共享。而科学数据共享政策是科学数据开放共享环境建设的重要内容, 可以从制度、标准和机制上保证科学数据的有效利用和发挥价值。为深入研究和了解各类主体在科学数据共享方面制定的政策, 本刊策划组织了“科学数据共享政策研究”专题。其中,《面向利益相关者的科学数据共享政策分析》一文从整体上归纳了各利益相关者制定的数据政策, 其他几篇论文则针对BBSRC、Dryad、PLoSOne、PUBMED进行了深度的调研和分析, 分析了科研基金资助机构、数据仓储、学术期刊、研究型大学图书馆在科学数据收集、加工、利用、服务整个数据管理与共享生命周期中的政策内容, 以期对今后我国各类机构制定相应的数据共享政策提供参考。

——编者

## 面向利益相关者的科学数据共享政策分析

屈宝强 王 凯 彭 洁 翟娟华

(中国科学技术信息研究所, 北京 100038)

**摘要:** 科学数据共享政策在规范共享过程中的利益关系与共享行为, 维护共享活动正常秩序方面有重要作用。文章主要对科研基金资助机构、期刊 & 出版商、数据存储机构等制定的数据政策进行分析, 提出今后要着力构建科学数据共享政策体系、明确政策核心内容、加强政策执行监督、完善科学数据标准体系等。

**关键词:** 科学数据; 数据共享; 数据共享政策; 数据仓储

中图分类号: G350

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2015.06.007

### Scientific Data Sharing Solution For Participating Subjects

Qu Baoqiang, Wang Kai, Peng Jie, Zhai Juanhua

(Institute of scientific & technical information of china, Beijing 100038)

**Abstract:** Scientific data sharing policy has important function in regulating the interests of all participants and maintaining the normal order of the open scientific data activities. This paper analyzes the data sharing policy of research funder, journal & publisher, data storage institutions, then puts forward some advices such as constructing scientific data sharing policy system, clarifying policy core content, strengthening policy

**作者简介:** 屈宝强\* (1980-), 男, 管理学博士, 中国科学技术信息研究所副研究员, 研究方向: 数据共享; 王凯 (1990-), 男, 硕士研究生, 研究方向: 数据共享; 彭洁 (1965-), 女, 中国科学技术信息研究所研究员, 研究方向: 信息资源管理; 翟娟华 (1978-), 女, 中国科学技术信息研究所馆员, 研究方向: 科技管理。

**收稿时间:** 2015年11月18日。

implementation supervision, improving scientific data standard system, etc..

**Keywords:** scientific data, data sharing, policy, data repository

## 1 引言

当前,数据日益成为科学研究的重要基础,成为现代科技基础设施不可缺少的组成部分。科学数据共享有助于提升科学数据利用效率,发掘科学数据价值。随着共享活动的深入发展,国际社会认识到科学数据共享是一项复杂的活动,不仅需要强烈的共享意识、坚实的技术基础,而且需要法规政策等来约束与规范共享行为。

科学数据共享领域的政策制定主要围绕共享活动所涉及的各个环节展开,平衡科学数据的知识产权和公众对科学数据应用的权利<sup>[1]</sup>。该领域的政策制定主体包括国际组织及政府机构、基金资助机构、研究机构、高校图书馆及数据仓储、期刊出版商等,他们从不同角度对共享活动中当事各方之间的利益关系与行为进行调整约束,维护共享活动的正常秩序,共同推进科学数据共享的良好实践。

Smith等将数据共享系统中的主要角色分为三大类:数据共享对象object(数据共享的单位),指任意粒度和类型的数据;数据共享主体actor,包括数据提供者Providers和数据消费者

Consumers;数据共享促进者,包括提供多种增值服务的分发者Distributors以及建立数据标准、从事工具研发的促进者Facilitators等<sup>[2]</sup>。

数据提供者由正式和非正式的、使数据容易被别人获取的组织和个人组成。有时数据提供者可能是数据生产者,数据也可以被存放在数据仓储、数据中心或档案馆,由这些机构承担传播数据的责任。期刊也可以是数据提供者,可以通过发表数据论文或通过提供支撑论文研究结论的补充材料或数据来促进科学数据共享。数据消费者(用户)是获取数据并在自己的工作中使用这些数据的个人和群体,数据生产者也可以是数据用户。数据共享促进者是数据共享过程中的重要利益相关者,主要是帮助数据消费者快速获得数据资源的相关机构和组织,例如科研基金资助机构、期刊&出版商、图书馆和数据中心等数据管理机构、研究机构等,各类数据共享促进者的相关政策对于科学数据的有效利用至关重要。

## 2 科学数据共享政策的相关研究

国外对科学数据共享政策体系的研究,主要包括:(1)介绍不同利益相关者制定的政策。例如

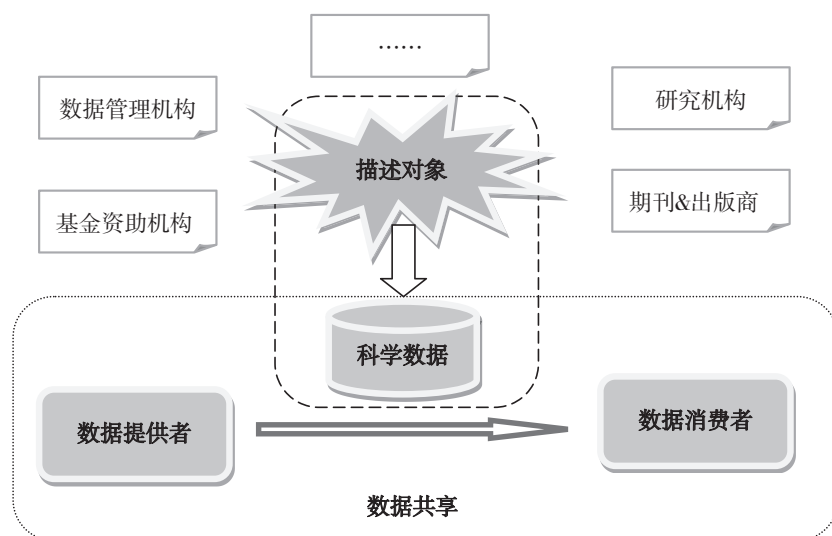


图1 数据共享体系中的各类参与主体

Socha总结了国际科学组织、基金资助机构、研究机构、学术图书馆、出版商和期刊的数据引用政策和制度<sup>[3]</sup>。Jubb分析了基金资助机构针对资助对象在数据管理、保存、发布及共享方面的政策要求<sup>[4]</sup>。Groenewegen等介绍了澳大利亚5所大学的机构数据管理与共享政策<sup>[5]</sup>。Savage等指出越来越多的期刊制定了数据共享政策，要求论文作者以某种方式与其他研究人员进行数据共享<sup>[6]</sup>。

(2) 不同层次政策之间的影响。例如北德克萨斯大学发起的DataRes项目，分析了2011–2012年间美国NSF推行数据管理计划对一些科研机构相关政策的影响<sup>[7]</sup>。Faniel等指出目前国际性数据共享政策非常复杂，研究人员普遍趋向于遵守本地区或本机构的共享政策<sup>[8]</sup>。

(3) 相关政策对科研人员的影响。Piwowar等利用文献计量学方法发现科研资助机构的数据共享政策和研究人员的数据共享行为之间有着显著的正相关性<sup>[9]</sup>。Daniel调查相关期刊中论文作者的数据共享意愿，发现即使该期刊有对应的数据管理与共享政策，但论文作者的共享意愿也并不是很强烈<sup>[10]</sup>。

国内对科学数据共享政策体系的研究，主要包括：(1) 国外相关政策法规的介绍。唐义等根据政策产生效力范围的不同，把目前国际上的科学数据共享政策法规体系分为宏观、中观和微观3个层次<sup>[11]</sup>。司莉等梳理了2005年以来国外有关科学数据管理与共享研究热点<sup>[12]</sup>。李娟等研究了CODATA、OECD两个国际组织共享原则的内容、特点、应用以及美国共享政策法规体系的构成及特点<sup>[13]</sup>。顾立平等对学术期刊、数据期刊、科研基金资助机构、科研教育机构等的政策进行了调研<sup>[14–17]</sup>。

(2) 国内相关政策法规的发展情况。例如：路鹏等介绍了我国的《中华人民共和国科学数据共享条例（建议稿）》<sup>[18]</sup>；王卷乐等对973计划资源环境领域项目数据汇交计划中的各环节进行了分析总结<sup>[19]</sup>；杨从科对我国农业科学数据共享发展中的政策问题进行了分析<sup>[20]</sup>。

(3) 政策法规的内容研究。例如：裴雷构建了上下位政策概念的一致性指标测算框架，讨论了我国科学数据共享政策在吸收、扩散和创新过程中的政

策文本质量，并提出改进建议<sup>[21]</sup>；黄如花等提出要加强我国科学数据管理的政策法规体系建设，完善与科学数据管理相关的规章制度，增强现行项目数据管理规定的操作性<sup>[22]</sup>；刘润达等探讨了我国科学数据共享的政策法规体系<sup>[23]</sup>。

### 3 各利益相关者制定政策内容分析

#### 3.1 科研基金资助机构

科研基金资助机构要求数据共享已经成为共识。美国国立卫生研究院（NIH）要求从2003年10月1日开始，所有向NIH申请项目经费在50万美元以上的科研人员必须提交一份数据共享计划或者数据保密说明。美国国家科学基金会（NSF）于2010年1月发布的项目管理指南规定：从2011年1月18日开始，所有提交到NSF的项目申请书必须包含一份不超过两页的“数据管理计划”附件，详细描述申请者如何管理和传播研究项目所产生的数据，包括：数据的类型、数据采用的标准、数据获取与共享政策、数据存档与保存计划。英国研究理事会（RCUK）制定了《数据政策的共同原则》，该原则为各个研究理事会成员的数据政策制定的总体框架。英国Wellcome Trust在2010年8月发布了《数据管理和共享政策》，要求所有申请者在研究建议书阶段就应考虑管理和共享数据的方法，必须在Wellcome Trust做出是否资助决定前提交一份数据管理和共享计划。加拿大卫生研究院（CIHR）要求研究人员在公布研究成果后把生物信息学相应数据存放到适当的公共数据库，并保留原始数据集至少五年。澳大利亚研究理事会Australian Research Council（ARC）要求，从2014年2月起研究人员需要在申请国家竞争性计划资助项目过程中，描述如何管理ARC资助研究所产生的研究数据。

这类主体制定数据政策的主要内容及特点包括：(1) 强调公共财政资金投入所产生的科学数据必须对公众开放。(2) 大多数机构对科研活动各个阶段的数据政策有详细规定，特别是要求在申请者提交资助申请时必须提供相应的数据管理规划（Data Management Plan），并在项

目实施阶段遵照执行,将此作为是否资助的重要条件之一<sup>[24]</sup>。(3)根据对研究主体知识产权保护的要求,一般数据开放都有一定的滞后期(Embargo)。

### 3.2 期刊&出版商数据政策

当前,许多国际重要的学术期刊都已经制定与数据开放、数据共享和数据利用的相关政策。例如传统学术期刊《BMC Evolutionary Biology》《PLoS One》《BMJ》《Science》《ZooKeys》等,其政策内容主要包括:(1)支撑研究论文主要研究结论的科学数据应以各种形式开放供读者查阅,以便重现研究过程,验证研究结果。(2)提交科学数据的方式主要包括将数据交给期刊存储;将数据存储于开放获取的综合性数据仓储(如Dryad)、领域数据仓储、机构知识库等;作者自存储数据;在读者提出相关数据的查阅和利用需求时,作者直接提供给读者。(3)在相应使用许可条件下,这些科学数据可以获取、传播和重复使用。(4)是否提交相应数据作为学术论文发表的重要前提之一。

最近几年出现了一批以发表数据论文为主要目的的数据期刊如《Scientific data》《Geoscience Data Journal》《Earth System Science Data》等,数据论文会对数据情况、数据采集和处理过程及数据文件格式进行说明,但不包含任何研究方法说明或研究结论,读者能了解数据的具体内容以及数据是何时、何地、如何收集的。一般在数据论文出版后,支撑论文结论的数据必须对读者开放,有的期刊还要求作者提供用于数据处理分析的工具,以满足重复性验证论文结论的需求。这类期刊的数据政策主要内容包括:(1)提供给作者关于该数据期刊的一套数据集描述工具、指南和模板,用来说明被接受论文的类型以及数据论文的结构与格式。(2)在数据存储要求方面,一般要求将数据集存放在期刊推荐的“可信数据仓储”。(3)支持数据引用方面,期刊不仅要求对数据论文本身进行引用,而且要求作者在参考数据时引用数据。(4)数据质量控制方法,要求通过同行评议对数据质量进行评议,或者将数据公开

接受使用者质量评估。(5)期刊一般提供推荐的唯一标识符(如DOI)以便将数据论文与存放在数据仓储的科学数据关联。

在期刊科学数据的管理政策框架方面,Dryad于2011年提出联合数据存档政策(Joint Data Archiving Policy, JDAP),目前已有包括《Science》《The American Naturalist》《Heredity》《Molecular Ecology》《The Journal of Evolutionary Biology》等在内的重要期刊、图书馆、出版社采用。

### 3.3 数据存储机构

高等院校的科学数据一般由图书馆负责管理。例如美国新墨西哥大学2009年8月启动DataONE项目,构建地球观测数据的基础设施。DataONE制定了3个方面的共享数据政策<sup>[25]</sup>:数据隐私与保密、知识产权、数据使用许可。

另外,还有一些独立的数据存储机构,例如Dryad、PANGAEA数据仓储等,在其数据政策方面,PANGAEA的数据管理政策遵循ICSU的政策和OECD的相关原则,提交到PANGAEA进行存储的数据,都遵循Creative Commons许可使用协议<sup>[26]</sup>。

我国西部数据中心制定了《科学数据与资料集成和共享管理条例》,详细说明了数据收集、管理、共享和长期保存的方式,同时制定《数据服务人员工作细则》对服务的质量和时效进行控制。复旦大学社会科学数据研究中心先后制定了平台元数据规范、数据监护规范和数据管理条例。国家地震科学数据共享中心制定了《地震科学数据共享管理办法》及4部配套实施细则,为地震科学数据共享的全面展开提供了制度保障。

这类主体数据政策的主要内容及特点包括:(1)明确指出数据仓储的基本任务是为相关机构研究人员提供数据存储服务,帮助其进行数据管理,保证研究人员科研活动的顺利开展。(2)其政策内容包括规定科学数据收录范围、存储要求、存储规范、存储数据类型、存储地点、权利声明等。(3)强调对所存储数据的开放利用及在利用过程中对数据生产者的权利保护,包括数据所有权管理、开放共享、许可协议、引用规范等。

### 3.4 其他相关机构

除上述各类利益相关者制定的数据共享政策外，还包括一些国家政府部门和国际组织制定的政策，例如欧盟在2006年12月颁布的《开放获取公共资助研究数据的原则和指南》报告中明确提出了13条原则与指导方针以促进公共资助科学数据的低成本、高效率获取。2013年2月，美国科学技术政策办公室（OSTP）发布了题为《增强政府资助科学研究成果的获取》的文件，要求联邦政府资助的研发项目所产生的科学论文和科学数据在发表后一年内通过机构知识库向公众免费开放。

从微观层面讲，每个研究机构是数据管理的重要主体。一些重要科研机构非常重视科学数据共享，例如美国疾病控制与预防中心（CDC）2000年就开始实施执行《数据发布与共享政策》<sup>[27]</sup>。另外，根据科研资助机构的要求，各机构也制定了相应的数据管理政策。例如：英美一些主要研究型大学例如普度大学、哈佛大学、加州大学、南安普顿大学、曼彻斯特大学、巴斯大学等已经开发了他们机构关于数据共享、数据管理和数据利用许可方面的政策。这些机构数据政策的核心是提供本机构研究人员数据存储规范，帮助其进行数据管理，详细规定科研活动中数据采集、加工、管理、分析、利用等环节的具体要求，并注重机构内以及机构之间在数据管理方面的协作。

## 4 思考与建议

通过上述分析可以发现，当前在科学数据共享政策方面已经取得了重要进展，包括科研基金资助机构、期刊&出版商、数据存储机构、研究机构在内的各个利益相关者已经有了良好的实践。但从整体上来看，还存在以下问题：一是政策内容方面，对科学数据共享中的数据标识、数据质量控制、数据长期保存方案等方面的规定杂乱、不统一，不利于科学数据的关联集成与再利用。二是在政策执行及监督方面，许多政策内容非常完善，但是执行效果不理想，没有切实与科研诚信、科技奖励、科技评价等方面的制度协同

推进。三是在政策延展性方面，政策制定局限在科学数据领域，还需要进一步考虑科研活动的状况，例如考虑到在同一领域使用不同科学仪器产生数据的格式标准化和数据交换问题等，可以尝试在开展科学数据引用著录时增加使用仪器型号方面的著录项。因此，我国今后科学数据政策的发展应考虑以下几个问题。

### （1）构建科学数据共享政策体系

在科研过程中每一类独立的利益相关者政策对科学数据共享的约束力有限，因此需要统筹设计，构建系统的、科学的数据共享政策体系，形成面向各类主体的政策框架和指南，引导和推进科研资助机构、期刊、出版社、数据中心、机构知识库等多个主体出台相互衔接、互为补充的政策。同时考虑现有政策状况，与《政府信息公开条例》《国家科技报告管理办法》《中国科学院关于公共资助科研项目发表的论文实行开放获取的政策声明》和《国家自然科学基金委员会关于受资助项目科研论文实行开放获取的政策声明》《科学数据共享条例》《国家科技计划项目科学数据汇交办法》《科学数据共享工程管理办法》《科学数据共享工程试点遴选和检查评估办法》和《科学数据分类分级共享及其发布策略》等相关政策协同推进。

### （2）明确科学数据共享政策核心内容

明确各个利益主体制定科学数据共享政策主要诉求和核心内容，例如根据本文分析，科研基金资助机构政策应加强“Data Management Plan, DMP”这一内容，强调在科研活动实施之前数据管理活动的预先筹划；期刊&出版商数据政策围绕“Data reuse”，强调科研活动中产生的重要成果——科学数据的再利用；数据存储机构政策围绕“Data curation”，强调数据管理的细节和过程。

### （3）加强科学数据共享政策执行监督

在现实中，对于科学数据共享政策的执行效果不是很理想<sup>[23]</sup>，需要根据各个国家、地区、机构的实际情况，建立切实的政策实施细则。一方面提升研究人员数据开放意识，加强数据集成和融合基础上的创新活动组织；另一方面从科研经

费投入、科研成果发布、科技评奖、职称职务晋升等方面体现对数据共享的要求。

#### (4) 完善科学数据标准体系建设

面向科学数据的管理和利用,建立各个环节数据处理的标准体系,包括科学数据唯一标识体系标准、面向各学科领域的科学数据元数据标准、科学数据引用著录标准、科学数据交换标准、科学数据分类体系等。

#### 参考文献

- [1] 朱雪忠,徐先东.浅析我国科学数据共享与知识产权保护的冲突与协调[J].管理学报,2007(4):477-482.
- [2] Smith K, Seligman L, Swarup V. Everybody Share: The Challenge of Data-Sharing Systems[J]. Computer, 2008, 41(9):54-61
- [3] Yvonne M Socha. Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data[J]. Data Science Journal, 2013, 12(13): 1-76.
- [4] Michael Jubb. UK Research Funders' Policies for the Management of Information Outputs[J]. The International Journal of Digital Curation, 2007(2):29-48
- [5] David Groenewegen, Andrew Treloar. Adding Value by Taking a National and Institutional Approach to Research Data: The ANDS Experience[J]. The International Journal of Digital Curation, 2013, 2(8):89-98
- [6] Savage C J, Vickers A J. Empirical Study of Data Sharing by Authors Publishing in PLoS Journals[J]. PLoS One, 2009, 4(9): e7078.
- [7] Martin Halbert. The Problematic Future of Research Data Management: Challenges, Opportunities and Emerging Patterns Identified by the DataRes Project [J]. The International Journal of Digital Curation, 2013, 2(8):111-122.
- [8] Faniel I M, Zimmerman A. Beyond the Data Deluge: A Research Agenda for Large-Scale Data Sharing and Reuse[J]. International Journal of Digital Curation, 2011, 6(1):58-69.
- [9] Piwowar H A, Chapman W W. A Review of Journal Policies for Sharing Research Data[C]. International Conference on Electronic Publishing 2008, Toronto: KMDI, 2008: 253-265.
- [10] Daniel D Reidpath, Pascale A Allotey. Data Sharing in Medical Research—An Empirical Investigation[J]. Bioethics, 2001, 15(2):125-134.
- [11] 唐义, 张晓蒙, 郑燃. 国际科学数据共享政策法规体系: Linked Science 制度基础[J]. 图书情报知识, 2013(3):67-73.
- [12] 司莉, 庄晓喆, 王思敏, 等. 2005年以来国外科学数据管理与共享研究进展与启示[J]. 国家图书馆学报, 2013, 22(3):40-49.
- [13] 李娟, 刘德洪, 江洪. 国际科学数据共享原则和政策研究[J]. 图书情报工作, 2008(12):77-80.
- [14] 吴蓉, 顾立平, 刘晶晶. 国外学术期刊数据政策的调研与分析[J]. 图书情报工作, 2015(7): 99-105.
- [15] 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(6): 53-60.
- [16] 杨云秀, 顾立平, 张瑶, 等. 国外科研教育机构数据政策的调研与分析——以英国10所高校为例[J]. 图书情报工作, 2015(5): 53-59.
- [17] 刘晶晶, 顾立平. 数据期刊的政策调研与分析——以 Scientific Data 为例[J]. 中国科技期刊研究, 2015(4): 331-339.
- [18] 路鹏, 苗良田, 莫纪宏, 等. 关于起草《中华人民共和国科学数据共享条例(建议稿)》的说明[J]. 国际地震动态, 2008(8): 41-46.
- [19] 王卷乐, 孙九林. 地球系统科学数据共享标准规范体系研究与应用[J]. 地理科学进展, 2009 (6):839-847
- [20] 杨从科. 中国农业科学数据资源建设研究[D]. 北京: 中国农业科学院, 2007.
- [21] 裴雷. 我国科学数据共享政策概念一致性与政策质量评估[J]. 情报理论与实践, 2013 (9):28-31.
- [22] 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.
- [23] 刘润达, 彭洁. 我国科学数据共享政策法规建设现状与展望[J]. 科技管理研究, 2010(13):40-43.
- [24] 朱艳华, 胡良霖, 袁雅琴. 国内外科研资助机构科学数据共享政策分析[J]. 中国科技资源导刊, 2015 (3):50-57
- [25] DataONE. Sharing Data: Legal and Policy Considerations [EB/OL]. [2014-08-05]. <https://www.dataone.org/best-practices/sharing-data-legal-and-policy-considerations>.
- [26] 中国科学院文献情报中心. 开放数据调研报告[EB/OL]. [2014-07-25]. <http://open-resources.las.ac.cn/drupal/?q=node/3064>.
- [27] National Center for Health Statistics. Centers for Disease Control and Prevention[EB/OL]. [2014-08-01]. <http://www.cdc.gov/DataStatistics/>.