



开放科学  
(资源服务)  
标识码  
(OSID)

# 科学数据引用：前提、挑战与实现机制

桂婕<sup>1</sup> 杨波<sup>2</sup> 屈宝强<sup>1</sup> 刘蔚<sup>1</sup> 陈晓文<sup>1</sup>

1. 中国科学技术信息研究所 北京 100038;
2. 北京科技审评中心 北京 101117

**摘要:** [目的/意义] 研究科学数据引用的重要前提、现实挑战与实现机制, 对于在科研活动中推广科学数据引用、保护科学数据知识产权具有重要意义。[方法/过程] 基于全生命周期管理、利益相关者分析等理论, 分析科学数据引用的发生条件、影响因素等, 开展对比与差距分析, 提出相关工作建议。[结果/结论] 科学数据引用要求把科学数据作为重要产出, 并记录其在科研过程中利用情况。当前科学数据引用实施中存在引用不清晰、溯源不稳定、关联不准确、追踪不便捷等问题, 科学数据引用实现依赖于各利益相关者协同推进政策体系、著录标准、技术基础设施建设, 当前我国需重点推广引用标准、探索科学数据评价、强化基础设施建设。

**关键词:** 科学数据; 数据引用; 著录标准

**中图分类号:** G301

## Scientific Data Citation: Preconditions, Challenges, and Implementation Mechanisms

GUI Jie<sup>1</sup> YANG Bo<sup>2</sup> QU Baoqiang<sup>1</sup> LIU Wei<sup>1</sup> CHEN Xiaowen<sup>1</sup>

1. Institute of scientific and technical information of China, Beijing 100038, China;
2. Center of science and technology review and assess, Beijing 101117, China

**Abstract:** [Objective/Significance] To study the important premises, realistic challenges and realization mechanisms of scientific data citation is of great significance for promoting scientific data citation and protecting intellectual property rights of scientific data in scientific research activities. [Methods/Processes] Based on the theories of life cycle management and stakeholder

**基金项目** 中央级公益性科研院所基本科研业务项目“中信所统一数据底座建设与运营管理”(ZD2023-04); 国家社科基金项目“科学数据引用的实现机制研究”(15BTQ053)。

**作者简介** 桂婕(1976-), 博士, 副研究员, 研究方向为数据治理、科技情报分析; 杨波(1979-), 硕士, 中级会计师, 研究方向为科技管理; 屈宝强(1980-), 博士, 研究员, 研究方向为科学数据、情报分析, E-mail: qubq@istic.ac.cn。刘蔚(1986-), 博士, 副研究员, 研究方向为数据挖掘、情报分析。陈晓文(1994-), 硕士, 助理研究员, 研究方向为信息资源管理。

**引用格式** 桂婕, 杨波, 屈宝强, 等. 科学数据引用: 前提、挑战与实现机制[J]. 情报工程, 2023, 9(4): 79-88.

analysis, this paper analyzes the occurrence conditions and influencing factors of scientific data citation, carries out comparison and gap analysis, and puts forward suggestions for promoting scientific data citation. [Results/Conclusions] Scientific data citation requires that scientific data be considered as first-class scientific output and that its use be documented in the scientific research process. At present, there are some problems in the implementation of scientific data citation, such as unclear citation, unstable traceability, inaccurate correlation, and inconvenient tracking. The realization of scientific data citation relies on the coordination of various stakeholders to promote the construction of policy system, description standards, and technical infrastructure. At present, China needs to focus on promoting citation standards, exploring scientific data evaluation, and strengthening infrastructure construction.

**Keyword:** Scientific Data; Data Citation; Description Standards

## 引言

引用是表达对他人的科研成果利用和致谢的重要方式之一。科学数据引用 (scientific data citation) 指通过一定的标识技术和机制, 对所使用的科学数据资源进行描述, 标识数据的来源, 通过科学数据引用在一定程度上促进了科学数据的知识产权保护, 便于科学数据使用状况的追踪与统计分析。

学者 Dodd 于 1979 年就提出把数据看作智力作品, 以与美国书目参考文献标准一致的方式描述和引用数据集<sup>[1]</sup>。21 世纪以来, 随着数据在科研活动中的重要性不断增加, 对科学数据引用的重要性认识不断提升, 许多机构在推进科学数据引用方面做了大量工作。在科学数据唯一标识体系方面, 2009 年成立 Datacite, 将提供科学数据的 DOI 注册、管理与应用作为一项重要服务。在科学数据引用索引体系构建方面, 2012 年, 汤森路透旗下知识产权与科技事业部正式推出科学数据引文索引 (DCI) 数据库, 提供用户工具来搜索和发现与研究有关的研究数据。在科学数据引用原则方

面, “FORCE11” 于 2014 年发布了 “FORCE11 数据引用原则共同宣言”, 提出了数据引用的八个原则: 重要性、荣誉和署名、证据、唯一标识、可获取、永久性、专指性和可验证性、互操作性和灵活性<sup>[2]</sup>。在科学数据出版方面, 2012 年 5 月, 自然出版集团推出数据期刊《科学数据》。在科学数据引用著录方面, 许多国际组织、数据中心和研究人员提出了不同的引用标准<sup>[3-6]</sup>, 引用涉及的主要元素包括: 作者、标题、出版日期、版本、出版商、标识符、位置等, 我国也于 2018 年 1 月发布科学数据引用标准。在实现科学数据引用的关键技术和体系方面, 自动生成引用和创建人机可读的引用<sup>[7]</sup>、使用时间戳查询<sup>[8]</sup> 和版本控制方法<sup>[9]</sup> 等相关方法开展动态科学数据引用方面都进行了相关探索。但是在学术交流实践当中, 依然存在科学数据引用意识不强、科学数据引用著录标准推广不够、科学数据引用基础设施不完备以及科学数据粒度和版本带来的技术问题等。本研究主要从理论上剖析科学数据引用的前提和动机, 分析存在的相关问题, 提出科学数据引用实现的相关措施。

## 1 科学数据引用的前提与动机

### 1.1 科学数据引用的前提

2014年1月荷兰莱顿举办的洛伦兹研讨会上首次提出FAIR原则，2016年3月Wilkinson等学者在《科学数据》(Scientific Data)上发表文章对它进行了系统性论述。FAIR原则即Findable(可查找)，Accessible(可访问)，Interoperable(可互操作)和Reusable(可重用)，要求在科学研究中产生的数据都需遵从这一原则。从科学数据生命周期管理的角度出发，科学数据引用发生在数据被研究者使用后，因此在研究过程中发现、获取并真正使用了这些数据，是引用产生的重要前提。

#### 1.1.1 科学数据应该被认为是重要的研究产出

在学术记录中，数据引用应与其他研究对象(如出版物)的引用同等重要。随着数据驱动科研范式的发展，数据越来越成为重要科研资源，同时科学数据作为研究过程中的重要产出形式得到了认可。例如，2011年在哈佛大学举办的IQSS会议有关数据引用的讨论中，明确提出引用数据“第一原理”：数据应被视为和文献同等重要的一流产出而实施引用，否则科研人员则不会在使用数据后主动地对其进行引用。

#### 1.1.2 科学数据可以被发现和获取

科学数据必须要存在于科研过程，只有科研人员能够比较容易地获取数据(有可能是自己制作)，才有可能在研究过程中使用并引用。如果数据不能被发现并获取，就谈不上对其引用。因此科学数据引用依赖于数据的出版发布，以及帮助科研人员发现数据

的基础设施，包括用于发现数据的索引、导航、门户等。

#### 1.1.3 科学数据的内容和价值应得到使用者的认可且被使用

真正满足研发过程需求的科学数据不仅依赖于数据本身的内容质量，更依赖于其对研究过程中的适应性、可开发性、可运算性。只有使用者认可了数据内容，并认为其可以在研究过程中起到相关作用，才会触发实施数据引用。实施数据引用要求科学数据被真正使用，为了某些计量学指标看上去很好的虚假引用是学术规范所不容许的，必须是在实际研究过程中支撑研究内容、对研究结论的形成有所帮助的数据，才能对其进行引用。

#### 1.1.4 使用和重用科学数据所产生的成果正式发布或者发表

使用和重用科学数据所产生的成果必须正式发布或者发表，且在发表的成果中以通用、可理解的格式表达出来才能真正实现科学数据引用，这些发表的成果是引用表达的场所，著录的文字表达是其表现形式。

### 1.2 科学数据引用的动机

科学数据引用的核心动力来自于对科研人员数据工作认可的需求、对科学研究过程再现的需求、对科学数据共享和利用效率提升的需求<sup>[10]</sup>。随着科学数据引用的发展，在数据管理、科技管理和情报研究领域，数据发现和影响力追踪也成了引用的重要动机。

#### 1.2.1 科学数据引用是承认数据工作者贡献的重要方式

随着科学研究活动协作的不断扩展，在科

研过程中需要涉及到科学数据工作相关人员的参与。这就需要记录不同类型参与者的贡献,以平衡合作过程中各利益相关者的需求和利益。

承认数据对相关研究产生贡献的方式可以有很多,包括致谢、引用、合著<sup>[11]</sup>。引用行为理论<sup>[12]</sup>通常认为,引用之前的作品是为了承认知识债务,提供知识的历史传承,并引导读者追本溯源。引用数据集的动力来自对在研究过程中产生的数据进行价值认可,同时数据也会存在于学术发展的历史传承中。对于研究人员来讲,这种在学术领域通过历史传承或者当前发布数据而得到尊重、职业发展和获得基金资助等是衡量其成功与否的重要内容。这导致学术群体内产生对规范性数据引用的期望,同时也引发通过各种方式对数据引用的宣传和培训。

### 1.2.2 科学数据引用是提升科学活动透明度的重要途径

引用数据提供了验证科研过程的途径,读者能够通过引用机制直接或间接获取用于论文的数据文件,把数据用于其他假设、建立模型或与其他数据相结合,对研究结果和过程的验证,从而使读者对论文进行评价。例如 *Applications in Plant Sciences* (APPS) 2016年8月签署 *The Center for Open Science's Transparency and Openness Promotion (TOP)*<sup>[13]</sup> 指南,提出促进数据透明、开放共享和可重现的8个标准,其中引用是第一标准。数据引用在提高科学透明度和结果再现性方面起着核心作用<sup>[14]</sup>,这在生物医学<sup>[15]</sup>、公共卫生<sup>[16]</sup>和生物学<sup>[17]</sup>等领域都得到体现。另外,在计算机科学领域,也有通过共享结果、代码及数据库研究脚本的方式实现研究过程和结果再现的尝试<sup>[18]</sup>。

### 1.2.3 科学数据引用是开展数据评价的重要依据

数据引用要求将更加完整的研究过程和结果提供给其他人,特别是通过适当方式将出版物与支撑其研究结论形成的数据链接<sup>[19]</sup>,数据引用著录内容从客观上可以提供必要的信息来解释数据的合法性,这可以作为是数据评价的重要基础。科学数据引用还可实现通过跟踪和溯源数据来间接进行数据质量评价,并且利用科学计量的方法对引用行为记录进行分析,通过计量结果间接评价数据质量(假设使用频率越高,说明该数据质量越高;数据存储依托的数据仓储权威度越高,说明该数据质量越高),还可以基于数据使用能力和水平开展科研人员、科研机构乃至区域数据能力评价。

### 1.2.4 科学数据引用是学术交流的重要内容

在表达形式方面,科学数据引用往往通过标准化形式体现引用行为和相关过程。通过研究人员、出版商、读者等认可的标准模板实现了科学数据引用创造、传播、组织和检索。科学数据引用则实现了科学共同体内对数据利用行为的认同<sup>[20]</sup>。在目前的科研管理体系当中,科学数据引用已经成为期刊出版商、资助机构和数据仓储的强制要求。科研人员遵从这些要求,主动或者被动地开展数据引用,将支撑论文主要观点的数据、论文使用的数据等内容清晰地著录,可以帮助科学数据在学术共同体内传播,也是体现学术素养和学术规范的重要内容。

## 1.3 科学数据引用实施面临的挑战

从科学数据本身来讲,由于数据的格式、类型、表现方式不同,数据引用面临诸多挑战。

### 1.3.1 科学数据多重粒度带来引用识别不清晰

在科学出版物中引用的数据可能是全部数据集或者子集，也可能是数据集不同部分的资源联合或不同数据集的组合。因此，数据被引用范围可以从单一数据记录到数据子集或数据集集合，形态和结构差异较大，这对科学数据引用实施会产生较大影响，也导致难以发现真正的数据引用行为。

### 1.3.2 科学数据多种来源带来引用溯源不稳定

科学数据引用需要明确科学数据来源，连同上下文信息一起来识别和发现数据，在此基础上开展数据溯源，掌握数据的原始状态，因此溯源模型对于将数据引用集成到数据管理流程具有较强的影响。由于科学数据来源的不确定，以及引用数据类型的多样化，造成了引用溯源不稳定，从而使得数据追踪和评价不准确。

### 1.3.3 科学数据多地存储带来引用关联不准确

科学数据引用机制实现的关键是数据有固定的存储位置，且能够长久存储。科学数据存储位置及时效性差异导致引用追踪、论文与数据关联不准确等问题。如果数据存储正式的、可信的数据仓储，则可以更好地保证可以通过永久标识关联或解析到数据内容；如果数据存储个人或者研究团队个人电脑的硬盘，则很难保证准确关联。

## 1.4 科学数据多样著录带来引用追踪不便捷

科学数据引用著录形式不统一，位置多样，可能在摘要、致谢和方法部分。例如在对某期刊的引用情况分析中发现，在 242 次科学数据引用中，出现在标题 4 次，摘要 59 次，正文 167 次，而参考文献仅为 12 次。有些引用著录

描述模糊，读者不清楚具体是什么数据（无数据名、无作者信息、无获取信息，无永久的唯一标识符），例如“以德温特创新索引作为数据源”“利用网页爬虫技术获取新浪微博数据”之类的表达<sup>[20]</sup>。这种多样化、模糊的引用著录造成了对同一文献中不同数据、不同文献中同一数据的引用追踪无从下手。

## 2 科学数据引用的实现机制

科学数据引用对于数据生产者、管理者、使用者等围绕数据价值实现的利益平衡具有重要作用，其实现依赖于各个利益相关者的共同努力。数据提供者主要是产生特定科学数据的个人、团体、组织，在科学数据引用实现过程中的主要职责包括产生科学数据，对科学数据进行加工整理，提升数据质量；形成科学数据描述性内容和元数据确保数据以可引用的格式存档，以合适的元数据提交数据；将元数据和数据提交数据仓储或者存储中心；在学术履历中列出其生产的数据集，将其作为一种学术成果。数据用户指使用科学数据的科研人员。他们应充分了解数据引用规范、遵守数据使用的各种约束，在使用过程中应对使用的科学数据进行合理规范著录，尽可能地使用数据集的唯一标识符（比如 DOI），并且通知数据提供者有关数据使用所形成的成果。在发表相关成果时，将数据作为出版物的附件一起提交或者建立出版物与数据之间的有效关联。数据分发者是数据生产者 and 使用者之间的中介，这些机构主要管理来自学者或机构提交的科学数据，参与建立统一的数据引用标准，支持并促进唯一

标识符使用,推动形成完整的数据出版体系,推动建设科学数据描述、发现、使用的基础设施和技术体系,使用自动化工具跟踪数据查看、搜索、下载、链接或引用行为,并反馈给数据生产和提供者。数据共享促进者包括相关国际组织、政府部门、基金资助者、学术共同体、公众。这些角色应积极参与制定数据共享政策、数据出版政策与引用规范,开发数据管理工具及支持性服务并且推动将科学数据纳入科研评价体系。

如本文第一部分指出的,上述各个利益相关者围绕数据引用标准制定、技术设施建设、政策制度出台开展了大量工作,但是科学数据

引用仍存在较大挑战。首先,从引用对象角度,需要确保科学数据大规模稳定存储、科学精细化管理、详尽细致性描述,永久长期性标识,把数据当作重要资产看待;其次,从引用行为角度,需要不断促进引用中的标准化著录、数据化追踪、指标化计算和真实性评价反馈,形成科学数据引用的运行体系;从引用实现环境角度,要求进行科学数据引用便捷性著录、评价方法针对性植入、技术工具全流程覆盖、引用意识和水平大幅度提升,将引用嵌入到科研活动过程中去。

基于上述考虑,本研究认为,科学数据引用的实现需要从以下几个方面推进:

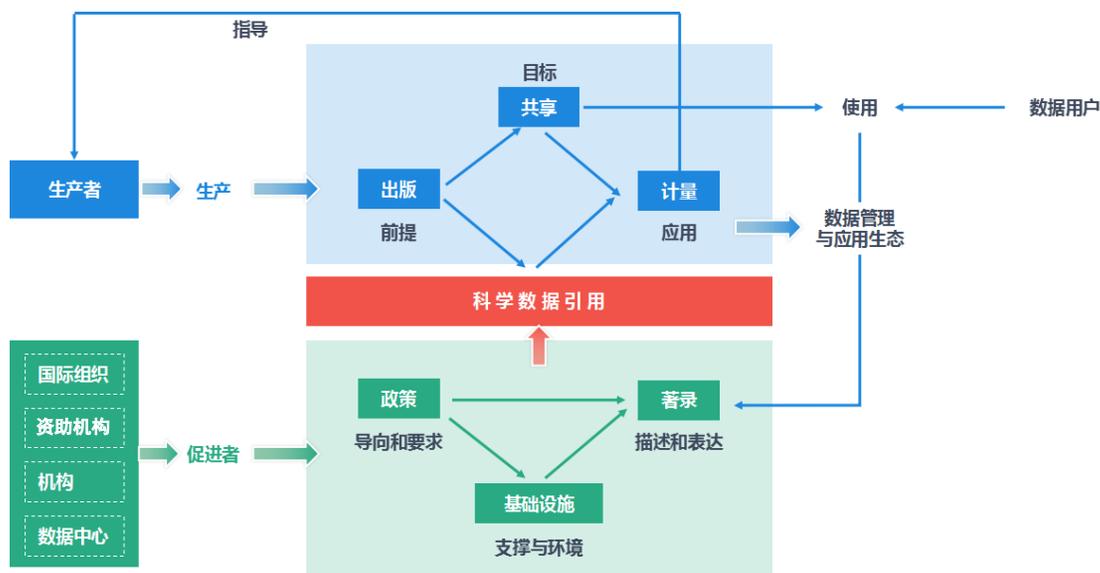


图1 科学数据引用的实现路径

## 2.1 科学数据共享、出版、引用、计量一体化发展

将科学数据引用融入科学数据生产制作、发布出版、共享利用、分析计量的过程,并且

全面准确记录与分析。从科学数据生命周期来看,数据被不同实体基于不同目的而生产、出版、传播、利用、评价、共享和重用,这些行为之间相互联系,相互作用,共同构成数据生命周期。

科学数据引用的出现和推广改变了传统的数据管理和使用生态体系，同时改变了学术交流的模式。

因此，需要在现有的生命周期中将引用作为重要环节，强调其对科学数据生产、出版、计量、共享的作用。数据引用可以被视为使用数据的信号，间接体现了数据工作者的能力和水平。数据计量可以测度和评价数据共享贡献度，促使数据引用走向规范化和标准化，方便了数据发现与获取，使得数据共享更可能发生。数据出版使得科学数据引用可以更好融入现有学术交流体系，允许更高质量数据以更加正式的方式被使用者发现，促进数据流转，从而产生数据引用。数据引用为数据计量提供了对象和内容，同时丰富和充实了科学计量相关方法和指标。

## 2.2 科学数据政策制度、著录标准、基础设施整体性建设

通过政策引导、基础设施保障、著录规则设定，不断完善在学术交流体系中科学数据引用行为、特征的规范化表达。政策指导层面提出数据引用的导向和要求，规范层面制定数据引用描述与表达的著录标准，基础设施层面建设支撑数据引用的工具及服务。

### 2.2.1 科学数据引用政策体系

在科学数据管理和共享过程中，涉及一些硬性的责权约束体系（例如法律政策、合同协议等）以及软性的约束体系（例如操作指南、伦理规范、学术要求等），通过这些可以实现科学数据共享中各个利益相关者的权益<sup>[21]</sup>，提出科学数据引用实施的内容、要求和方法。科

学数据引用需要从政策体系、共享协议和伦理规范角度分析各个利益方的责任与权益，并通过制定政策、编纂数据共享协议约束各方的利益关系与行为、规范数据用户和数据提供者之间的权利和义务，反思数据获取中的伦理问题。通过保障科学数据利用过程中各利益相关者权益，使得数据使用更加顺畅，进而保障数据引用的实现。

### 2.2.2 科学数据著录标准

科学数据引用需要对引用行为进行一致的形式化描述、规范化表达。各类组织、专业协会和期刊可以通过提供重要的数据引用标准促进规范化，同时提出这些表达的模板。由于各类标准化数据引用方法被鼓励并使用，以及长期稳定的数据存储和技术体系，使得数据使用者可以通过数据引用著录内容很容易查找和使用数据。

引用是学术作品交流传播的足迹，根据这些足迹可以判断数据流通和传播的方向，可以为使用/重用这些数据的用户进行画像，这些行为的分散和多样性可以为使用数据是否有序、有目的性提供线索。由引用著录所体现出客观存在的数据被不同科研人员使用的特点、模式和规律，可以不断充实和丰富科学计量的相关应用，基于数据引用著录内容开展分析可以帮助分析科研人员思想谱系，形成数据用户和数据提供者之间的有向网络，更好把握数据的利用情况。

### 2.2.3 科学数据引用基础设施

科学数据引用的基础设施用来支撑数据引用格式和标准形成、确认数据及引用的永久性、方便数据集检索和互操作、理解数据溯源

和贡献追踪。科学数据引用基础设施应该解决几个问题：第一，解决科学数据大规模存储和管理的问题，为引用实施中数据溯源和验证提供基础。主要通过数据仓储的专业数据监管工作来实现。第二，解决科学数据精细化描述与标识的问题，为引用过程中准确定位和导航提供基础。主要是通过科学数据唯一标识（DOI、ARK）及其解析系统来实现。第三，解决科学数据引用著录和生成的问题，为引用行为记录和自动分析提供基础。主要是通过统一著录标准以及相应的工具体系来实现。第四，解决科学数据识别和计量的问题，为引用行为挖掘和规律总结提供基础。主要是通过建设不同形式的科学数据引文索引库，以及以此为数据基础开展分析评价来实现。

### 2.3 科学数据引用利益相关者协同推进

科学数据引用要求数据生产者、数据使用者以及其他利益相关者各司其职，协同推进科学数据引用意识提升、标准执行、挖掘利用。从改变科学数据利益相关者关系角度来看，数据引用改变了传统的科学研究过程中的权益分配和职责范畴。表面上看，数据引用行为主要发生在数据的使用环节，但是在数据引用实施过程中，需要各个利益相关者在数据查询获取、分析处理、归档保存、交流重用等行为各个环节各司其职，共同努力<sup>[22]</sup>。没有所有利益相关者合作，不可能有效开展科学数据开放共享，建立引用机制，并进行数据影响力度量。

从参与主体角度来看，科学数据引用的实现依赖于各利益相关者对科学数据引用的认同，从而形成学术群体内遵守的通用行为准则，并

协同进行技术、标准、行为等各个方面的努力。

## 3 我国促进科学数据引用的相关建议

我国将科学数据引用作为推进科学数据共享与管理的重要手段，开展了一系列相关工作。在科学数据引用相关政策方面，2018年发布的《科学数据管理办法》中明确要求：科学数据使用者应遵守知识产权相关规定，在论文发表、专利申请、专著出版等工作中注明所使用和参考引用的科学数据。在科学数据著录标准方面，国家参考文献著录标准（GB/T 7714-2015）添加文献类型标识“DS”数据集，科学数据的引用纳入参考文献著录内容，同时2018年实施的科学数据引用标准（GB/T 35294-2017）规定了科学数据引用元素描述方法、引用元素详细说明、引用格式等方面的内容。在科学数据引用基础设施建设方面，已经形成“国家高能物理科学数据中心”等20个国家科学数据中心、“国家重要野生植物种质资源库”等30个国家生物种质与实验材料资源库，为可引用科学数据资源的长期存储和科学管理提供重要基础。围绕实现科学数据引用的重要路径，结合我国科学数据管理办法落实相关要求，为推动我国科学数据引用的发展，提出以下建议。

### 3.1 加强科学数据引用标准推广应用

鼓励各类基金资助机构、期刊出版社、数据中心形成促进科学数据引用的相关政策，引导和加强学者的数据引用意识。同时，加强科学数据引用的宣传推广，让每一个科研人员能

够认识到其重要性,营造整个科学领域“用则引”的氛围。在科研过程产生的不同类型、不同形式的产出中尽可能对引用行为开展著录,体现数据工作的重要性。可以考虑使用 GB/T 2015-7714 参考文献著录标准开展著录,利用“DS”标识数据引用。也可以以我国 2018 年 1 月发布的《信息技术科学数据引用》(GB/T 35294-2017) 标准为基础开展著录。

### 3.2 将科学数据贡献融入学术评价体系

破除“四唯”是我国深化科技体制改革激发创新活力的重要举措,当“破四唯”成为广泛共识,如何“立新标”成为重点,科学数据工作理应成为科研评价的重要内容。可以通过记录数据发现、浏览、使用等全过程,强化计量分析,为科学数据工作改善和价值发挥提供支撑。在此过程中,加强科学数据描述、检索、关联、发现、计量等理论方法的研究,开展数据共现、数据共引、数据聚类、数据网络分析,探索研究提出数据 H 指数、数据引用指数等,丰富科学评价内容。加强数据与各类科研产出之间的关联分析,以论文与数据之间的链接分析为基础,拓展数据与数据、数据与专利、数据与报告等其他产出关联分析的理论方法。

### 3.3 强化科学数据引用实现的基础设施网络体系

深化国家科学数据中心、领域数据中心、机构数据仓储的综合服务体系建设,既兼顾宏观上总体领域、区域布局的合理性,同时注重每个数据仓储在科学数据获取、存储、质量控制、安全等方面的系统性。通过元数据互联互

通,构建全国一体化科学数据服务网络,实现从单个数据存储机构(数据中心、机构知识库、共享平台等)服务到协同服务网络过渡。探索从其他科技信息资源管理系统的工作流中抽取数据,将科学数据与研究过程中的实体、产出、活动关联。

### 参考文献

- [1] DODD S A. Bibliographic references for numeric social science data files: Suggested guidelines[J]. Journal of the American Society for Information Science, 1979, 30(2): 77-82.
- [2] MARTONE M. Joint Declaration of Data Citation Principles[EB/OL]. [2016-08-18]. <https://www.force11.org/datacitationprinciples>.
- [3] ALTMAN M, KING G. A proposed standard for the scholarly citation of quantitative data[J]. D-lib Magazine, 2007, 13(3): 5.
- [4] LAWRENCE B N, JONES C M, MATTHEWS B M, et al. Data publication[EB/OL]. [2018-09-01]. <http://purl.org/oai/oai:epubs.cclrc.ac.uk:work/43641>.
- [5] GREEN T. We need publishing standards for datasets and data tables[J]. Learned Publishing, 2009, 22(4): 325-327.
- [6] STARR J, GASTL A. A metadata scheme for Datacite[J]. D-Lib Magazine, 2011, 17(1/2): 1-7.
- [7] SILVELLO G. A methodology for citing linked open data subsets[J]. D-Lib Magaz, 2015, 21(1/2): 6.
- [8] PROLL S, RAUBER A. Scalable data citation in dynamic, large databases: Model and reference implementation[C]. IEEE International Conference on Big Data, 2013: 307-312.
- [9] FERRO N, SILVELLO G. The Road Towards Reproducibility in Science: The Case of Data Citation[C]. 13th Italian Research Conference on Digital Libraries (IRCDL 2017), 2017: 20-31.
- [10] SILVELLO G. Theory and practice of data citation[J]. Association for Information Science and Technology,

- 2018, 69(1): 6-20.
- [11] FECHER B, FRIESIKE S, HEBING M. What Drives Academic Data Sharing?[J]. PLoS ONE, 2015, 10(2): e0118053.
- [12] NICOLAISEN J. Citation analysis[J]. Annual Review of Information Science & Technology, 2007, 41(1): 609-641.
- [13] STODDEN V. Enabling reproducible research: Open licensing for scientific innovation[J]. International Journal of Communications Law and an Introduction to the Joint Principles for Data Citation Policy, 2009, 13(13):1-25.
- [14] NOSEK B A, ALTER G, BANKS G C, et al. Promoting an open research culture: Author guidelines for journals could help to promote transparency, openness, and reproducibility[J]. Science, 2015, 348(6242): 1422-1425.
- [15] BAGGERLY K. Disclose all data in publications[J]. Nature, 2010, 467(7314): 401.
- [16] The Academy of Medical Science. Reproducibility and reliability of biomedical research: improving research practice[R/OL]. [2017-08-18]. <https://acmedsci.ac.uk/viewFile/56314e40aac61.pdf>.
- [17] CARR D, LITTLER K. Sharing Research Data to Improve Public Health: A Funder Perspective[J]. Journal of Empirical Research on Human Research Ethics, 2015, 10(3): 314-316.
- [18] BLOOM T, GANLEY E, WINKER M, et al. Data Access for the Open Access Literature: PLOS's Data Policy[J]. PLOS Biology, 2014, 12(2): e1001797.
- [19] FREIRE J, BONNET P, SHASHA D. Computational reproducibility: state-of-the-art, challenges, and database research opportunities[C]. AcmSigmod International Conference on Management of Data. ACM, 2012.
- [20] 孟阳. 科研数据计量分析框架与实证研究[D]. 北京: 中国科学技术信息研究所, 2017
- [21] FERRO N. Reproducibility Challenges in Information Retrieval Evaluation[J]. Journal of Data and Information Quality, 2017, 8(2): 1-8.
- [22] FERRO N, FUHR N, JAERVELIN K, et al. Increasing Reproducibility in IR: Findings from the Dagstuhl Seminar on "Reproducibility of Data-Oriented Experiments in e-Science"[J]. AcmSigir Forum, 2016, 50(1): 68-82.