

【信息管理】

# 数据管理视角下的 科研诚信管理困境与发展路径研究

夏义堃 钱锦琳

**【摘要】**[目的/意义]科研诚信是科技创新的基石,随着数据驱动型研究范式的普及推广,数据伪造、篡改等科研不端现象频发,凸显了科研诚信管理与科研数据管理之间紧密关联。[方法/过程]文章综合运用系统分析法、案例分析法、比较分析法,分析科研诚信管理与科研数据管理联动发展的交互强化作用、渐进累加作用,并以活动理论为指导从诚信管理主体、客体、规则、技术、共同体、分工六要素及其互动关系出发,探讨当前科研诚信管理困境。[结果/结论]文章建构了数据管理视角下科研诚信管理解释性框架,并提出科研诚信管理与数据管理联动共进的发展路径,包括以完备的数据制度推进诚信管理的整体布局,以精准的数据质量推动诚信管理的体系建构,以高效的数据开发深化诚信管理内生动力。

**【关键词】**科研数据;科研诚信管理;数据管理;科研不端

**【作者简介】**夏义堃(1967-),女,南京大学数据管理创新研究中心教授,博士生导师,南京大学数据智能与交叉创新实验室;钱锦琳(通讯作者)(1993-),女,南京大学数据智能与交叉创新实验室博士研究生(南京210023)。

**【原文出处】**《情报资料工作》(京),2023.6.15~24

**【基金项目】**本文系国家社会科学基金重大项目“基于数智融合的信息分析方法创新与应用研究”(批准号:22&ZD326)的研究成果之一。

科研诚信是科学研究的基础,对营造理性科研氛围、促进科研创新具有重要意义。近年来,英国、澳大利亚等多国政府、高校、出版机构相继颁布科研诚信政策以规范科研行为,我国先后于2009年、2018年出台科研诚信建设意见,2022年成立科研诚信与负责任创新专委会,从国家层面进一步加强了科研诚信管理的顶层设计。然而,囿于急功近利的科研氛围,国内外违背诚信要求的科研行为屡禁不止,特别是随着数据驱动型研究的普及,科研不端行为多表现为伪造数据、篡改数据等数据不端行为,揭示了数据管理失衡引发科研诚信管理缺位的内在关联。据此,本文以数据管理视角探析科研诚信管理困境与发展路径,首先从科研诚信管理与数据管理之间的联动机制入手,分析二者在管理方式、制度规则等

方面存在的适配关系,以此厘清数据管理与科研诚信管理互为促进的途径与作用方式,为梳理数字环境下科研诚信管理的主要内容与关键突破口奠定理论基础。其次,从数据是科研诚信的事实依据、数据行为是科研诚信的结果表征角度出发,借由活动理论,分析科研诚信管理活动的主体、分工与协作关系,以此阐释数据驱动型范式下科研诚信管理困境实则是科研诚信管理与数据管理联动的割裂。最后,从多主体数据管理行为与科研诚信管理理念、职责的有机结合出发,提出数据管理视角下科研诚信管理的解释性框架及其实施路径。

## 1 相关实践与研究回顾

### 1.1 科研诚信的内涵界定

科研诚信,也被称为科学诚信或学术诚信<sup>[1]</sup>。

2009年我国科学技术部、教育部等十部委联合颁布《关于加强我国科研诚信建设的意见》，从道德规范正面引导层面将其界定为“科研人员以追求真理、实事求是、崇尚创新、开放协作为精神指导，以法律法规和科学道德准则为行为指导，遵循科学共同体公认的行为规范”<sup>[1]</sup>。美国国立卫生研究院立足科研流程视角阐述科研诚信内涵，即遵循普遍认可的职业准则和规范，在提出、执行和评估科研的过程中，使用诚实和可验证的方法；在报告科研结果时，遵守规则和指南<sup>[2]</sup>。

### 1.2 科研诚信管理研究评述

近年来，国内外学界关于科研诚信的研究主要集中在科研诚信政策研究与科研诚信行为分析两大方面。在政策研究中，早期研究关注政策文本内容的解读，如科研诚信内涵<sup>[3]</sup>、参与者职责<sup>[4]</sup>。现阶段研究逐渐将诚信政策与诚信行为深度融合，如有学者发现科研诚信政策欠缺对个体诚信行为的关注<sup>[5]</sup>，就此提出制定具体学科科研诚信行为指南<sup>[6]</sup>；在科研诚信行为研究中，人员的性格特征<sup>[8]</sup>、所处的组织环境<sup>[9]</sup>以及利益获取<sup>[10]</sup>等因素皆对诚信行为产生影响。部分学者通过对比中外高校处置科研不端行为的办法，从开展诚信教育<sup>[11]</sup>、建立学术界信赖关系<sup>[12]</sup>以及利用数字技术加强审查力度<sup>[13]</sup>等方面提出我国处置科研不端行为的完善之路。随着数据驱动型研究范式的普及，科研诚信中的数据问题与数据行为日益受到关注，有学者将与数据有关的科研诚信问题称为数据诚信<sup>[14]</sup>。2017年欧盟宣布启用新版《欧洲科研诚信行为准则》，更加重视科研数据质量与管理，明确科研机构应承担的具体责任<sup>[15]</sup>。此外，学界也从数据管理的多元视角探讨科研诚信行为，如从研究者视角，人们发现通过数据分析手段伪造数据的行为比凭空捏造数据的行为更常见<sup>[16]</sup>；从期刊视角，严格的出版前质量控制手段有助于解决数据不端行为问题，如加强图像数据审核以确保论文可靠及研究结果可重现等<sup>[17]</sup>；从图书馆业务视角，学者提出可通过提供数据验证服务、发展数据馆员等途径打击科研数据造假现象<sup>[18]</sup>。可见，数据问题已成为当前科研不端现象的重要表征，强化科研数据管理已成为科研诚信管理的重要手段。

### 1.3 科研诚信管理中数据问题的现状表征与研究进展

科研诚信与科研不端看似相互对立，实则是一体两面，分别从正反两个视角、理想与现实两个层面反映了科研诚信管理对于科研创新乃至社会进步的重大影响。早在1986-1989年间，美国众议院将在科学研究及其传播、交流过程中，伪造、篡改、剽窃数据或实验步骤、歪曲作者身份的行为统称为“科研不端行为”<sup>[19]</sup>。我国政府部门分别从多个维度界定科研不端行为，例如科技部侧重项目管理过程中科技成果的真实性<sup>[20]</sup>；教育部重视以学术成果的规范性来判定科研行为的合规性<sup>[21]</sup>；国家自然科学基金委员会聚焦基金项目从申请到结题、成果发表与应用过程中的科研不端行为<sup>[22]</sup>。虽然三部委的侧重点各异，但从内容要件上均包含了数据伪造、剽窃、捏造等违反科研诚信的具体行为。换言之，科学研究中的数据问题及其数据管理本身已构成科研诚信管理的基础要件，加强科研数据管理是开展科研诚信管理的应有之意，任何带有主观故意的数据造假行为都是科研不端的重要表现，应置于科研诚信管理、科研数据管理的双重考量下加以解决。

在现实中，科研不端以林林总总的投机行为及其危害性警示人们科研诚信管理与科研数据管理的重要性。美国科研诚信办公室的一项调查发现，41%的科研不端行为涉及数据问题，部分科学家承认至少有过一次伪造、篡改、捏造数据的行为<sup>[23]</sup>。近几年来，数据行为失范已然成为科研不端事件的突出表现，部分学者从撤稿论文的视角反诉科研不端行为<sup>[24]</sup>，发现科研数据管理失当与科研不端事件发生之间有着紧密的对应关系，是科研诚信管理失衡的典型表现。也有学者从相关学科撤稿论文的调查中发现，因数据问题引发的撤稿事件占全部撤稿事件的31%<sup>[25]</sup>，生命科学领域中因操纵数据、篡改图像数据等行为引发的撤稿事件占比高达39%<sup>[26]</sup>。尽管已有学者尝试论证科研数据存储与科研诚信建设之间的逻辑关联与作用机制，提出科研不端行为的应对策略<sup>[27]</sup>，但鲜有研究从科研数据管理与科研诚信管理的内在关联及其作用机理角度展开深层次的分析。据此，本文拟从数据管理视角剖析科研诚信

管理的结构框架与运行特征,揭示二者间互为依存、相互促进的联动机制,以此寻求科研诚信管理的应对方案。

## 2 科研诚信管理与科研数据管理的联动机制

数据是科学研究的基本构成要件,其获取、利用、传播与科学研究、学术出版的全过程管理密切相关;同时,数据也是科研诚信管理的基本工具,实验记录、临床数据、图像资料等既是科研不端行为调查的重要证据,也是科研过程是否合规、是否运用了可靠而有效的研究方法、是否符合数据伦理的第一手验证材料。随着数据密集型研究的普及,基于数据分析与挖掘、数据可视化、数据交流的数据探索已经成为现代科学研究的重要组成,非客观的数据收集<sup>[28]</sup>、不公正的数据分析<sup>[29]</sup>、利益导向的数据传播<sup>[30]</sup>等引发的科研不端行为大幅度增加。相较直接剽窃、捏造,隐藏在数据加工处理中的数据篡改、图像操纵、选择性数据利用等行为往往不易被发现,被视为“隐性学术不端”。大量事实表明,科研诚信管理活动与数据管理活动之间在价值理念、发展周期、管理方式、制度规则等方面存在梯度适配关系,数据管理理念、方法、规则不仅构成科研诚信管理的重要依托,也直接影响诚信管理的总体状况与建设成效。

### 2.1 交互强化机制

交互强化机制系指科研诚信管理与科研数据管理在管理方式、运作机制中的相互渗透、优势互补与彼此增强,从而形成 $1+1>2$ 的科研管理效能。尽管二者的目标、对象与管理框架分属不同的逻辑进路,科研数据管理聚焦科研数据采集、加工、存储、传播等微观层面数据管护的“有形管理”,侧重于科研探索过程与结果的数据化展示,而科研诚信管理则强调科研机构、人员等主体在科学精神与伦理道德层面的价值追求与科研自觉,如诚信意识、科研自律乃至违规惩戒等均属于宏观层面的“无形约束”,但借助于数据的开发利用,能够实现科研数据“有形管理”与科研诚信“无形约束”的有机结合,从而通过数据行为的规范重塑并体系化诚信管理逻辑。

一方面,贯穿科研数据生命周期的规范化数据管理可以延展和诠释科研诚信的精神意蕴,通过数据采集、建模、挖掘、利用、归档等关键环节的标准化

管理,能够强化科研诚信的事前管控与科研人员的行为自觉,实现诚信管理对科学研究过程的全覆盖;另一方面,建构在科研诚信基础上的科研数据管理,以贯通科学精神与数据诚信为主旨,在目标追求、治理框架与主体关系等方面都与科研诚信管理高度契合,有助于开放获取、数据共享、FAIR原则等主导性数据管理理念的确立,从而提升科研过程与结果的透明度、推进数据的再利用与理论、方法的创新。

### 2.2 渐进累加机制

渐进累加机制系指科研诚信管理与科研数据管理需要通过缓慢的投入积累与持续的管理强化形成强有力的技术、制度、平台、数据等科研基础设施,一旦一方形成规范、得到普及和认同,则会带动另一方的调整优化、递进优化,进而又推动自身迈向更高水准的发展阶段,形成彼此呼应、迭代共进的累加协同效应。

2010年世界科研诚信大会确立的《新加坡声明》将包括数据在内的科研资源的良好管理纳入科研诚信管理的基本原则<sup>[31]</sup>。数据存储中心、数据平台等基础设施建设及其数据管理规范化水平的持续增强,使得科研数据共享和学术不端惩治同处于具有高度复合性、对应性数据关系的网络结构中。由于“当代数据科学的变革性特征不仅源于大数据和强大计算的可用性,还源于数据作为研究材料和证据来源的概念化根本转变”<sup>[32]</sup>,围绕数据可发现、可访问、互操作与重用所建构的数据管理体系不仅有助于增强图像、数据等科研资源的保存、共享能力,为科研不端惩治提供充分的一手证据,还进一步规范了科研数据采集访问、分析加工、共享开放等业务流程,推动了科研数据链、创新链前后工序的业务衔接与相互监督,客观上对数据不端行为形成有效约束。同时,科研诚信制度的不断完善、科研不端行为识别和查处技术的不断提升,在强化科研数据痕迹管理、证据功能的同时,也同步促进了数据链上下游管理标准、技术和政策的逐步健全,开放存取、开放数据、FAIR原则的普及以及资助机构、出版商等多主体数据管理规则的实施,均蕴含了开放、诚实、公正等科研诚信原则,并通过技术、方法、流程以及监管方式等数据管理水平的迭代升级,形成全新的科

研诚信管理模式。

### 3 数据管理视角下科研诚信管理困境

#### 3.1 基于数据驱动的科研诚信管理活动分析框架

科研诚信管理与科研数据管理之间紧密的内在联动表明,数据密集型研究范式下,真实性、完整性、准确性等科研数据管理的质量诉求包含于科研诚信管理所追求的目标原则之中,并构成检验科研不端行为的指导性依据,科研诚信管理在很大程度上决定了科研数据管理的本质规定与活动形态。据此,本文引入活动理论,拟从科研数据管理活动的视角分析当前科研诚信管理现状与问题。

活动理论起源于德国古典哲学、马克思主义辩证法与苏联心理学,列昂捷夫将活动要素进行细化分解,构成了活动—行为—操作层级的基本结构<sup>[33]</sup>,在此基础上,芬兰学者构建了基于主体、客体、共同体、工具、规则、劳动分工六要素的活动理论模型<sup>[34]</sup>。从活动理论出发,科研诚信管理是科研人员及其团队成员以及政府、出版机构等利益相关者在科研诚信这一价值导向下,对包括数据问题在内的所有诚信管理任务的协作与分工,不可避免地涉及诚信主体、客体、共同体、工具、规则、劳动分工六要素,并形成科研诚信管理的活动要素框架,以此揭示科研诚信管理主客体间关系,理清诚信管理中的数据问题“谁来管”“管什么”“怎么管”,进而剖析数据诚信问题的“动机”与“条件”及相应的“目标”管理行为(见图1)。

为(见图1)。

在数据管理情境下,科研诚信管理的“主体”指发起科研活动的个体或组织,即科研人员及其科研团队;“客体”指主体行为对象,即数据行为;而与主体共享客体要素的其他活动参与者统称为“共同体”,泛指政府、高校、科研院所、资助机构、期刊出版机构以及科研同行、社会公众等;“工具”指主体为实现科研目的而使用的数据平台、软件等基础设施;“劳动分工”指科研主体以及共同体在应对数据问题相关科研诚信管理任务的具体分工;“规则”则为协调主客体关系的规章制度。“动机”指促使主体展开诚信管理行为的导火索,“条件”泛指技术、政策等各项外部资源,一旦“动机”与“条件”形成适配和发酵,则会产生相应的活动指向“目标”。同时,结合列昂捷夫对活动类型的划分,可将科研诚信管理活动分解为目的性活动、功能性行为、常规性操作组成的三层递进式层级结构<sup>[33]</sup>。当下数据行为不规范事件的大规模发生以及人们对开放透明科研环境需求的双重影响,促使科研主体聚焦科研诚信中的数据管理问题表征及其背后成因,并将诚信管理活动同步解构为针对数据现象的具体活动,如数据收集、分析等具体数据操作中的动机、条件与行为选择等。

从国内外已有数据问题引发的科研不端事件分析来看,科研诚信管理的主体、客体、规则、共同体等构成要素、内在运行与科研数据管理问题有着紧密

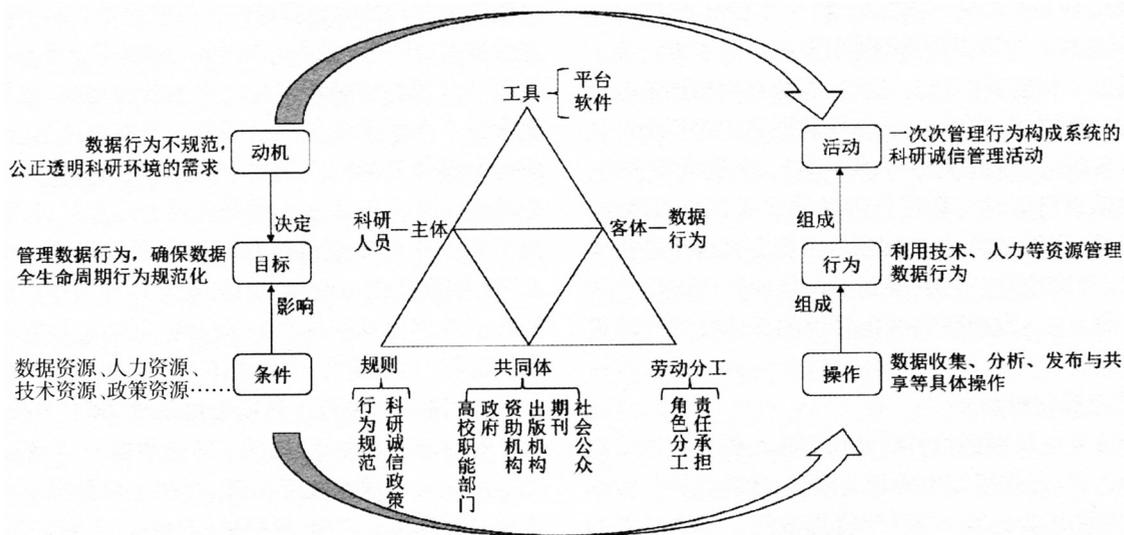


图1 科研诚信管理活动要素框架

的对应关系,二者不仅在活动的作用对象、动机目标、方法手段上具有一致性,还在活动的时间维度上具有交互同步性,人们的数据理念、数据方法与数据行为不仅直接嵌入科研活动过程,还有机融入科研诚信管理活动各要素,在一定程度上已成为科研诚信管理状态的风向标。

### 3.2 科研诚信管理困境分析

通过对国内外数据造假等科研不端活动的深入剖析,发现当前科研诚信管理所面临的问题是多方面的,既涉及主体观念认知、共同体职责协同,也涉及科研诚信管理和数据管理的技术、规则等多方面问题。

(1)诚信认知偏差与数据行为失范的共谋。科学研究本是科研数据与学术规范、诚信伦理等高度融合、互动发展的产物,但当偏重于新知识发现与理论、方法和技术创新的科研逻辑与侧重于结果绩效和任务指标的市场逻辑发生冲突时,由数据、图像等构成的论文、著作就具有了通向终身教职、科研奖项、人才头衔的可兑换性,容易导致诚信认知的偏离。有学者指出:“科研论文因欺诈和不当行为而被收回数量的不断上升,反映了科研活动的主要目的——对知识无私追求——的扭曲和边缘化。”<sup>[35]</sup>一方面,经济理性带来科研动机与目标的功利化倾向,使得数据、资料等科研资源通过学术发表环节逆向转化为人们所需要的科研资本、资源,进而对科研主体的诚信认知及其行为选择提出考验。当利益诉求凌驾于科研诚信与科学精神之上时,篡改数据、操纵图像等科研不端行为在所难免。例如,受名利驱使,黄禹锡授意他人更改实验步骤以获取理想数据的行为<sup>[36]</sup>,本质上是诚信认知扭曲下的数据不端;另一方面,数据标准、数据存储等业务管理规则的不健全也为激励、机会和压力交织下的数据投机行为打开了方便之门。简言之,诚信认知偏差易导致诚信意念薄弱及其数据行为失范,而数据不端行为又将动摇主体科研信念,并造成诚信管理+数据管理错综交织的恶性循环。

(2)诚信管理协同性不足与共同体数据管理缺位的叠加。科研不端事件的惩治、诚信理念及其技术支持的普及等是共同体相互协调、共同作用的结果,

意味着科研主体及其参与者的适当协调、合理分工能够提供相应支持,并刺激数据等科研资源开发。但囿于科研管理的部门分割与部门利益,诚信管理主体与共同体之间、不同共同体成员之间对话基础薄弱,尚未形成合力。例如哈佛大学在成立科研不端调查委员会时,委员会成员之间配合度低且未尽职责,导致事后调查出的涉嫌数据造假的论文数量远高于委员会的实际调查结果<sup>[37]</sup>。同时,政府主管部门、项目资助机构、科研院所以及期刊出版部门、数据中心等并未在科研数据管理制度、标准实施中形成衔接紧密的科研数据合作网络。加之“目前的科学激励体系可能过分强调创新,而不利于验证”<sup>[38]</sup>,数据获取、核验、交换、共享等业务往往成为科研活动的边缘性工作,进一步激化了主体与共同体成员间诚信管理、数据管理协同性不足的矛盾,进而导致数据造假等不端事件的持续性涌现。如荷兰心理学家 Diederik 通过伪造数据成功发表了 50 余篇论文,其间期刊编辑与同行评议专家等并未严格履行数据审核职责,导致数据管理未能形成闭环<sup>[39]</sup>。日本学者的研究发现,部分审稿专家利用职务之便,迫使投稿人引用自己的数据,使投稿人成为“不诚实的从众”<sup>[40]</sup>,均反映了数据管理规则缺位下数据沦为工具性、功利性筹码的现实困境。

(3)诚信管理技术开发滞后与数据管理基础薄弱的累积。有学者指出:“如果说过去的科研欺诈是手工的,那么新的数字工具几乎将其变成了一个常规或自动化的过程”<sup>[41]</sup>。图像处理软件、数据可视化软件等数据工具的广泛使用,也为操纵图形、有选择地挖掘数据和复制其他作者的文本等科研不端行为的更新换代提供了便利,真实折射出“工具”要素在促进数据管理与科研不端惩治中融合管理所发挥的重要作用。首先,现有诚信管理在方法上多停留在诚信委员会有限的专家调查与定性分析层面,难以针对科研不端事件形成完整的数据画像,图像抄袭、数据剽窃等自动检测技术、方法的工具性局限客观上增强了数据伪造等不端行为的隐匿性,助长了部分科研人员“走捷径”心理。有学者基于 20000 篇论文图像复制的调查指出:“过去十年中,有问题图像的论文流行率显著上升”<sup>[42]</sup>,如小保方晴子通过计算机

技术篡改图像数据骗取期刊认同,将成果在《自然》期刊发表<sup>[43]</sup>。其次,科研数据标准、数据存储平台及其数据应用工具的不足也极大制约了科研诚信管理工作的开展。美国贝尔实验室以科学严谨性闻名,而舍恩却以数据存储技术的缺乏为借口,坚称原始数据已丢失,增加了诚信调查的难度<sup>[44]</sup>,数据存储能力以及团队科研管理的不足为舍恩的不端行为创造了可乘之机。可见,数据密集型科研由于数据种类、规模与利用等特定要求的提升而进一步增强了诚信管理和数据管理技术复杂性要求,需要以数据平台、存储中心等坚实的工具性支撑为基础。

(4)诚信管理制度泛化与数据管理规范不足的交织。纵览国内外现有科研诚信行为准则等政策规范,大多属于宏观指导型和原则阐述型,如我国《关于进一步加强科研诚信建设的若干意见》《欧洲科研诚信行为准则》《澳大利亚负责任研究行为准则》等均明确了科研诚信管理的内涵、原则与任务要求、组织机构等,具有较强的宣教引导性,而各类不端行为处理办法则侧重于科研不端行为概念及其查处流程、惩治方式等规定,主要从撤销学术成果、业内警告等角度展开,多属于事后管控。然而,有效的诚信管理包括了从研究产出过程到研究成果转化的科研全生命周期,其制度设计包括数据行为规范理应覆盖科研主体、共同体,对资助机构、科研院所、研究人员、期刊出版部门的科研规范和诚信管理义务予以明确。同时,还应契合不同学科诚信行为要求、不端表现的特殊性,需要在数据政策的针对性、标准性与深度上加以完善,以确保科研全过程、全主体、全领

域数据诚信管理的有法可依。当然,“仅仅意识到负责任的数据管理是科研诚信的基础是不够的,还需要在数据管理规划、文件组织与命名、存储与备份、元数据和文档以及数据注册和引用等方面进一步规范”<sup>[45]</sup>,以规制人们的数据行为,对数据准备和分析中无限制的自由裁量权予以必要的约束。

综上,当前科研诚信管理的困境并非由单一要素所致,而是在数据驱动型科学研究过程中,人们的数据管理方式、价值实现渠道以及数据原则规范未能实现与科研诚信管理工具、规则、分工等活动要素的相互衔接(见图2),表现为科研数据管理与诚信管理联动的割裂,并反映在“活动泛化—行为失范—操作失当”中,从而造成科研诚信管理与数据管理原则规范的分离、价值实现驱动力的偏移和管理方式的封闭。究其原因,则源于科研诚信与数据管理规则的不健全、科研共同体行动协同不足、职责分工模糊、科研主体诚信意识薄弱、诚信及数据管理工具供给有限进而表现为客体—论文、资料等科研记录的不准确、不完整、不真实等一系列症结。

#### 4 科研诚信管理与数据管理联动共进的发展路径

2000年,美国科研诚信办公室将“数据获取、管理、分享以及所有权”排在科研诚信教育九大核心内容的首位<sup>[46]</sup>。2007年,第一届世界科研诚信会议集中讨论的一项重要议题就是科研成果出版及科研记录的监测管理<sup>[46]</sup>,并在2008年发布了题为“促进科学和科学记录诚信”的声明(ICSU CFRS)<sup>[47]</sup>。同年,美国卫生研究院进一步提出:“好的科学需要有良好的科研记录,以增进科学研究的责任和诚信”<sup>[48]</sup>。不难看出

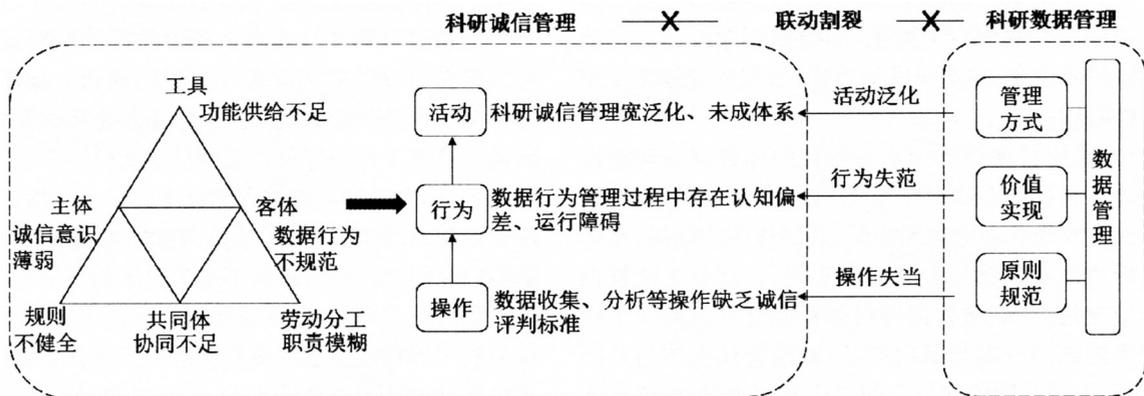


图2 科研数据管理视角下科研诚信管理困境分析

出,科研诚信管理与记录保存等科研数据管理的有机衔接逐渐成为学界共识,现代科研诚信管理越来越多地依赖于数字、图像等数据资源,科研诚信管理与数据管理是科研主体的共同责任,多主体数据行为、数据关系与科研诚信管理理念、职责的有机结合正在塑造现代科研诚信生态,并构成数据管理视角下科研诚信管理的解释性框架(如图3)。也就是说,基于数据在诚信管理各环节间的有序流转,能够通过数据制度的完备、数据质量的监管、数据开发的合规来改进和提升科研诚信管理的观念认知、管理方法、技术基础与制度规范,从而促进诚信管理活动主体、客体、共同体、规则、工具、劳动分工要素的激活与优化,以实现数据价值从诚信共识到共创的跃变。

#### 4.1 以完备的数据制度推进诚信管理的整体布局

数据驱动型研究导向下,科研诚信管理的整体布局既需要在诚信制度中明确数据诚信的重要意义与治理原则,不断提升数据管理与诚信管理互为促进、协同发展的理念共识,还要以专门的科研数据管理法规为主干,以与科研诚信管理相关的数据采集获取政策、记录存储政策、分析利用政策、共享交换政策、安全监管政策等规范性文件及其相配套的数据标准、技术参数、流程规则为依托,形成涵盖科研诚信管理全流程、全方位的数据管理制度体系。为确保科研诚信管理中数据问题解决的有法可依、有法必依和数据不端行为的违法必究、执法必严,还需要从以下方面增进科研主体、共同体的诚信管理力量。

(1)提升数据共识。数据理念作为诚信活动的指引,直接影响科研主体、共同体的数据行为选择与对象结果。总体而言,科研主体对于开放科学、数据共享等科研数据管理法规政策的认知并不理想。欧洲大学协会(EUA)在2020-2021的年调查中发现,虽然55%~70%的高校认识到科学数据管理、数据公平获取与共享的重要作用,但在实践中高度重视并组织实施的只有15%~25%<sup>[49]</sup>。同时,个体科研人员的数据管理意识仍需强化,2022年的《欧洲科研数据图景》报告指出:超过三分之二的受访者要么没有听说过FAIR原则,要么没有完全理解其含义<sup>[50]</sup>,表明科研主体的数据意识、数据管理能力难以适应数据管理与诚信管理的建设需要。

(2)营造数据合作网络。诚信管理与数据管理的联动推进在科研人员、政府主管部门、科研资助机构、科研院所、期刊出版部门、图书馆、数据中心等不同利益相关者之间构成了复杂的多维关系依赖型网络。面对诚信管理、数据管理之间既有的交集空间和独立的业务领域、层次类型,不仅科研人员等科研主体需要明确其数据管理职责与诚信自律义务,资助机构、学术期刊、数据平台等科研共同体也需要在各自的业务活动中明确相应的管理边界和服务规则,从而形成多主体责任分明、衔接紧密的可操作性数据管理制度体系。如澳大利亚研究理事会通过加强与学术共同体、政治共同体等利益相关者团体之间的合作,形成了良好的高校科研治理外部生态,以此加强科研人员对科研数据存储、利用等数据行为

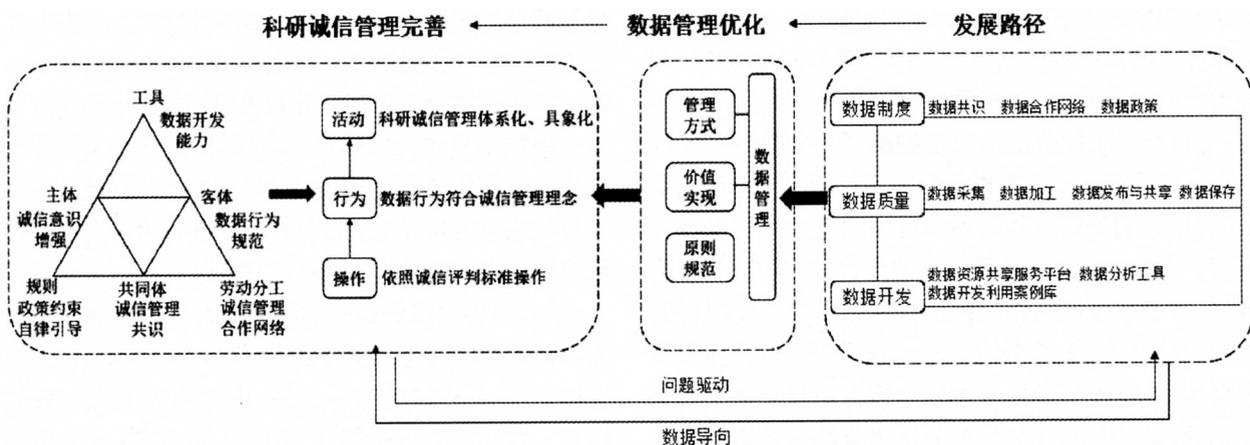


图3 数据管理视角下科研诚信管理解释性框架

的管理,并从同行评议、科研评价等角度提高科研人员对科研诚信的认知<sup>[51]</sup>。

(3)科学制定数据政策。科研数据管理政策的制定既关系到国家战略性科研数据资源的保护和科研数据共享交换与知识传播的推进,也关系到科研诚信管理任务的落地实施、科研不端事件的监控惩治。在一定意义上,也是对科研人员、团队与相关机构的数据行为、数据权力进行必要的约束。由于科研诚信管理贯穿于科学研究与出版传播的全过程,因而需要建立起覆盖科研、出版全生命周期的数据管理制度体系,如科研数据管理计划、元数据规范、数据开放共享规则、数据存储要求、数据审计规程等。同时,应注重制度的强制约束性与科研主体、共同体道德自律性的协调平衡,加强学会性数据公约、诚信守则引导性功能的发挥,从而形成门类齐全、结构合理的科研数据管理法规政策体系。

#### 4.2 以精准的数据质量推动诚信管理的体系建构

有学者认为,相比较而言,科研成果的“发表较为容易,但数据才是真正的野兽”<sup>[52]</sup>。数据质量的高低一定程度上反映了科研主体的诚信状态与数据理念、管理水平。数字化时代科研诚信管理目标的实现,必须高度依赖科研数据的采集、加工与共享、利用,并对数据质量管理提出更高的要求,既要确保数据操作免受非学术因素的干扰,又能以真实、准确、及时、完整等质量指标反向保证各数据流程操作均符合FAIR原则,并成为评判数据可信的基本依据。

(1)数据采集阶段的质量控制。科研数据采集过程中诚信管理所面临的质量问题不仅决定了科研成果走向,还影响到后续科研不端行为的发生与后果严重程度。主要的数据问题既有伪造数据、捏造数据、编造数据、带有明显倾向性的挑选数据以及故意使用有偏差的实验设计。在生物、医学、心理学等领域还涉及特定群体、特定人员数据采集的知情告知与未遵守人类受试者保护等相应问题,需要从规范数据采集的业务流程、加强实验室记录管理、完善元数据标准以及普及数据伦理、隐私保护等入手。

(2)数据加工阶段的质量控制。数据挖掘、分析、计算、可视化等活动不仅是数据价值提升的关键,也是科研不端行为频发的高风险阶段。所涉及的数据

不端类型多样,诸如故意的计算错误、误导性数据解释、选择性数据加工、运用软件工具进行图像/数据伪造、剽窃等。有必要将数据管理作为科研诚信预警的第一道防线,不断细化科研主体数据管理职责,加强团队成员数据业务衔接和相互检查,提高设计与分析透明度,完善实验室工作日志,加大数据管理软件工具开发,开展数据管理内部审核评估,以及时发现数据不端问题。

(3)数据发布与共享阶段的质量控制。该阶段是数据价值逐渐由科研个体向群体传递和转换的开端,面临数据封锁、数据结果不可再现、特定人员隐私侵害等问题,涉及期刊编辑、资助机构、科研院所、学会、同行专家多个主体。对此,应鼓励科研数据开放共享,建立严格的数据汇交制度和多主体数据审核机制。目前,国外已有部分生物医学期刊通过设置数据诚信分析(data integrity analysis)专职审核员岗位,严格把控数据的真实性<sup>[53]</sup>。此外,还应在遵守国际出版伦理委员会(COPE)核心实践基础上,完善期刊出版的数据管理政策。

(4)数据保存阶段的质量控制。图像、数据以及文档资料保存为科研不端行为的查处提供了必要的证据支撑,但科研数据保存必须建立在规范的数据管理基础上。已有研究表明,数据通常会在论文发表两年后丢失,并且可获取概率呈现逐年递减趋势<sup>[54]</sup>,所面临的主要问题是科研数据丢失以及保存不当。对此,英国工程和物理科学研究理事会不仅对科研院所提出数据共享责任,还要求其在数据生命周期管理基础上至少提供十年的数据保存和管理服务<sup>[55]</sup>。同时,还应加强数据基础设施建设,建立多学科数据存储中心,并加大监管力度,确保数据安全。

#### 4.3 以高效的数据开发深化诚信管理的内生动力

数据是科研诚信的真实表现,是科研不端的惩治依据。数字化环境下,无论是科研诚信教育,还是科研人员与成果可信度提升以及科研不端行为的发现查处均需要以充分的科研数据开发为前提。

(1)完善数据资源共享服务平台建设。高水平数据存储库/数据中心以及数据资源共享服务平台建设,有助于打破科研数据个人保管、内部共享、部门封锁的困境。在丰富数据采集、汇聚的总量、种类,

推动跨部门、跨领域数据共享开放的同时,有助于科研过程与数据管理透明度的提升,从而吸引更多科研人员和机构参与到科研过程体验与结果验证中,客观上强化了科研诚信的过程监管,并化解困扰已久的“可复现性危机”,起到诚信预警作用。

(2)加强数据分析工具开发。数据分析工具的开发和分析能力的快速提升,能够在科研数据的高质量汇聚,算法、模型和软件工具的开发中增强科研不端文献检测系统的功能,包括对数据实体、元数据、说明文档等进行统一的规范化表述,进而识别形式各异的数据异常,及早发现分散在不同数据源中的数据/图像伪造、捏造等现象。例如,运用区块链的P2P组网技术、密码学算法等特征,根据科研数据的具体类型,建立全局范围的公共区块链、区域性的私有区块链等,为研究人员、高校机构、期刊等提供接口,从研究立项开始,跟踪科研全过程。如此,有助于开展科研不端行为的轨迹画像,归纳不同学科、不同类型主体不端行为特征,以开展针对性诚信管理,减少科研投机的可乘之机。

(3)加强数据利用案例库建设。良好的数据管理案例以其严格的流程管理、标准化的操作、协同化的组织而成为诚信管理的典型案例,对于促进科技创新和科研诚信建设都有着重要而长远的意义。除正面典型案例外,负面数据管理失当案例,其警示作用同样重要。如英国研究与创新署通过成立中心记录库,在组织内部及组织之间形成数据不端行为案例的共享机制,以此推进不同组织联合开展更为全面的诚信管理监督<sup>[56]</sup>。

## 5 结语

本文通过分析数据管理与科研诚信管理之间潜在的联动依存关系,进一步明确数字环境下科研诚信管理本质上是一项以数据和数据行为作为管理对象并覆盖科研全过程的管理活动,进而深入剖析数据管理作用于科研诚信管理的途径方式。同时,基于活动理论挖掘诚信管理不同要素及其互动关系所形成的管理困境,分析得出主体诚信意识模糊、客体数据行为不真实、工具供给不充分等一系列问题的症结在于数据管理与诚信管理联动割裂。据此本文提出数据管理视角下加强科研诚信管理的解释性框

架,有必要通过数据管理制度、行为、技术的强化,营造正向的科研诚信环境和开放数据生态,以提升科研数据利用的效能,加速数据向知识、智慧的科研创新转化。

## 参考文献:

- [1]科学技术部科研诚信建设办公室. 科研诚信知识读本[M/OL]. 北京: 科学技术文献出版社, 2009: 7[2022-08-12]. <https://www.jxyty.cn/10363/pdf/16245955587971932ba4f7ef5aec.pdf>.
- [2]国家自然科学基金委员会. 关于加强我国科研诚信建设的意见[EB/OL].[2022-06-10]. <https://www.nsf.gov.cn/publish/portal0/tab442/info62169.htm>.
- [3]What is Research Integrity[EB/OL].[2022-06-10]. [https://grants.nih.gov/policy/research\\_integrity/what-is.htm](https://grants.nih.gov/policy/research_integrity/what-is.htm).
- [4]Steneck N H. Fostering integrity in research: definitions, current knowledge, and future directions[J]. *Science and Engineering Ethics*, 2006, 12(1): 53-74.
- [5]Marcel C L. The politics of research misconduct: congressional oversight, universities, and science[J]. *The Journal of Higher Education*, 1994, 65(3): 261-285.
- [6]刘宇, 伍丹炜, 叶继元. 什么是科研诚信? 基于政策文本的话语分析[J]. *江南大学学报(人文社会科学版)*, 2022, 21(3): 35-45.
- [7]Shannon L L, Alexandra K, Joann M M, et al. An Update to Scientific Decision Making: The Academy of Nutrition and Dietetics' Scientific Integrity Principles[J/OL]. *Journal of the Academy of Nutrition and Dietetics*. [2023-01-18]. <https://www.eatrightpro.org/-/media/files/eatrightpro/practice/research/sip-updated2022.pdf>.
- [8]熊新正, 胡恩华, 修立军, 等. 科研诚信行为影响因素研究综述[J]. *科学管理研究*, 2012, 30(3): 39-42.
- [9]李真真, 黄小茹. 中国科研诚信面临的突出问题及解决路径[J]. *科学与社会*, 2017(3): 107-120.
- [10]Brienna M Larrick, Johanna T Dwyer, John W Erdman Jr, et al. An updated framework for industry funding of food and nutrition research: managing financial conflicts and scientific integrity[J]. *The Journal of Nutrition*, 2022, 152(8): 1812-1818.
- [11]冯菲, 袁子晗. 世界一流大学图书馆科研诚信教育的实践与启示[J]. *图书馆建设*, 2023(2): 150-159, 181.
- [12]曹南燕. 大学科研中的诚信问题[J]. *清华大学学报(哲学社会科学版)*, 2004(2): 5-9.
- [13]谢宇. 负责任研究行为及其实现进路[D]. 安徽: 中国科学技术大学, 2022: 124.

- [14]姚长青,田瑞强.开放科学中的数据诚信问题研究[J].科技与出版,2019(1):130-135.
- [15]中华人民共和国科学技术部.欧盟将启用新版《欧洲科研诚信行为准则》[EB/OL].[2022-06-10]. [https://www.most.gov.cn/gnwkjdt/201705/t20170516\\_132793.html](https://www.most.gov.cn/gnwkjdt/201705/t20170516_132793.html).
- [16]Gorman D M, Elkins A D, Lawley M. A systems approach to understanding and improving research integrity[J]. Science and Engineering Ethics, 2019, 25(1): 211-229.
- [17]Heck S, Bianchini F, Souren N Y, et al. Fake data, paper mills, and their authors: the International Journal of Cancer reacts to this threat to scientific integrity[J]. International Journal of Cancer, 2021, 149(3): 492-493.
- [18]黄如花,李楠.高校图书馆参与科学数据验证的前景分析[J].图书情报工作,2016,60(3):15-20.
- [19]Marcel C L. Research Misconduct[J/OL]. Society, 1994, 31: 6-10[2023-08-25]. <https://link.springer.com/content/pdf/10.1007/BF02693224.pdf>.
- [20]中华人民共和国科学技术部.国家科技计划实施中科研不端行为处理办法[EB/OL].[2022-06-16]. [https://www.most.gov.cn/xxgk/xinxifenlei/fdzdgnr/fgzc/bmgz/200811/t20081129\\_65718.html](https://www.most.gov.cn/xxgk/xinxifenlei/fdzdgnr/fgzc/bmgz/200811/t20081129_65718.html).
- [21]中华人民共和国教育部.高等学校预防与处理学术不端行为办法[EB/OL].[2022-06-16]. [https://www.moe.gov.cn/srcsite/A02/s5911/moe\\_621/201607/t20160718\\_272156.html](https://www.moe.gov.cn/srcsite/A02/s5911/moe_621/201607/t20160718_272156.html).
- [22]国家自然科学基金委员会.国家自然科学基金项目科研不端行为调查处理办法[EB/OL].[2022-06-16]. <https://www.nsf.gov.cn/publish/portal0/tab442/info79519.htm>.
- [23]Office of Research Integrity Annual Report 2002[EB/OL].[2022-09-13]. <http://ori.dhhs.gov/html/publications/annual-reports.asp>.
- [24]韩磊.2010-2019年中国生物医学论文的撤稿趋势及学术不端演变特征[J].中国科技期刊研究,2021,32(2):158-165.
- [25]易耀森.被撤销医学论文数据学术不端行为与防范对策研究[J].中国科技期刊研究,2020,31(3):276-280.
- [26]管茜,董克,夏义堃.基于撤稿论文的生命科学数据管理问题与对策研究——以基础生命科学领域为例[J].图书与情报,2021(3):47-56.
- [27]邢文明,陈继丽,王张华.面向科研诚信的科研数据管理保存:逻辑关联、作用机制与实现策略[J].图书情报知识,2021,38(6):134-143.
- [28]Ekbia H, Mattioli M, Kouper I, et al. Big data, bigger dilemmas: a critical review[J]. Journal of the Association for Information Science and Technology, 2015, 66(8): 1523-1545.
- [29]Resnik D B. Data fabrication and falsification and empiricist philosophy of science[J]. Science and Engineering Ethics, 2014(20): 423-431.
- [30]Ioannidis J P A, Munafò M R, Fusar-Poli P, et al. Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention[J]. Trends in Cognitive Sciences, 2014(18): 235-241.
- [31]Singapore Statement on Research Integrity[EB/OL].[2022-09-13]. <https://www.wcrif.org/guidance/singapore-statement>.
- [32]Sabina L. Data Governance is Key to Interpretation: Reconceptualizing Data in Data Science[J/OL].[2022-03-21]. <https://hdr.mitpress.mit.edu/pub/4ovhpe3v/release/6>.
- [33]Wilson T D. Activity theory and information seeking[J]. Annual Review of Information Science and Technology, 2008, 42(1): 119-161.
- [34]Liu D R, Shih M J. Hybrid-patent classification based on patent-network analysis[J]. Journal of the American Society for Information Science & Technology, 2011, 62(2): 246-256.
- [35]Dennis T, Russell C. Research misconduct in business and management studies: causes, consequences, and possible remedies[J]. Journal of Management Inquiry, 2020, 29(2): 174-187.
- [36]黄禹锡论文造假曾震惊韩国[EB/OL].[2022-08-15]. <https://xsgf.pku.edu.cn/alfx/gwal/1297193.htm>.
- [37]Kochan C A, Budd J M. The persistence of fraud in the literature: the darsee case[J]. Journal of the American Society for Information Science, 1992, 43(7): 488-493.
- [38]Nosek B A, Alter G, et al. Promoting an open research culture: author guidelines for journals to promote transparency, openness, and reproducibility[J]. Science, 2015, 348(6242): 1422-1425.
- [39]Martin E. Final report on Stapel also blames field as a whole[J]. Science, 2012, 338(6112): 1270-1271.
- [40]Shibayama S, Baba Y. Dishonest Conformity in Peer Review[J/OL].[2022-07-21]. <https://ssrn.com/abstract=2550008?>
- [41]Biagioli M, Lippman A. Introduction: Metrics and the New Ecologies of Academic Misconduct[EB/OL].[2022-08-15]. [http://direct.mit.edu/books/book/chapter-pdf/273263/9780262356565\\_caa.pdf](http://direct.mit.edu/books/book/chapter-pdf/273263/9780262356565_caa.pdf).
- [42]Bik E M, Casadevall A, Fang F C. The prevalence of inappropriate image duplication in biomedical research publications [J]. American Society for Microbiology, 2016, 7(3): 1-8.
- [43]小保方晴子论文造假事件[EB/OL].[2022-08-15]. <https://xsgf.pku.edu.cn/alfx/gwal/1297192.htm>.
- [44]王阳.重大科研不端案例与美国科研诚信制度的演变及其对中国的启示[J].自然辩证法通讯,2021,43(8):77-84.
- [45]Linda C G. Scientific integrity and ethical considerations for the research data life cycle[J]. Scientific Integrity and Ethics in the Geosciences, 2017(10), 133-156.

[46]Kritikos M. Safeguarding Research Integrity in Europe: An Object of Increasing Legal Attention[M]. Kritikos M. Finding Common Ground: Consensus in Research Ethics across the Social Sciences Advances in Research Ethics and Integrity. England: Emerald Publishing Limited, 2017: 199-210.

[47]Organisation for Economic Co-operation and Development. Best Practices for Ensuring Scientific Integrity and Preventing Misconduct[EB/OL].[2022-09-15]. <https://www.oecd.org/sti/inno/40188303.pdf>.

[48]National Institutes of Health Office of the Director. Guidelines for scientific record keeping in the intramural research program at the NIH[EB/OL].[2022-08-23]. [https://oir.nih.gov/sites/default/files/uploads/sourcebook/documents/ethical\\_conduct/guidelines-scientific\\_recordkeeping.pdf](https://oir.nih.gov/sites/default/files/uploads/sourcebook/documents/ethical_conduct/guidelines-scientific_recordkeeping.pdf).

[49]Rita M, Bregt S, Federica G, et al. From principles to practices: Open Science at Europe's universities[EB/OL].[2022-08-13]. <https://www.eua.eu/downloads/publications/2021%20os%20survey%20report.pdf>.

[50]European Research Data Landscape: Final Report[EB/OL].[2023-01-13]. <https://fair-impact.eu/news/europeanresearch-data-landscape-final-report>.

[51]Australian Research Council. Research Integrity Policy[EB/OL].[2023-03-20]. <https://www.arc.gov.au/about-arc/program-policies/research-security-and-integrity/research-integrity-policy>.

[52]Plutchak T S, Kaplan L. A library perspective: data wranglers in library land: finding opportunities in the changing policy landscape[J]. The Serials Librarian, 2016, 70(14): 14-25.

[53]Declan B. Researchers have finally created a tool to spot duplicated images across thousands of papers[J]. Nature, 2018, 555(7694): 18.

[54]Timothy H V, Arianne Y K A, Rose L A, et al. The availability of research data declines rapidly with article age[J]. Current Biology, 2014, 24(1): 94-97.

[55]Digital Curation Centre. Overview of Funders' Data Policies[EB/OL].[2022-07-23]. <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-datapolicies>.

[56]Research Councils UK. RCUK Policy and Guidelines on the Governance of Good Research Conduct Frequently Asked Questions[EB/OL].[2023-03-24]. <https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-RCUKPolicyGuidelinesGovernanceOfGoodResearchConductFAQsJAN18.pdf>.

## Research on the Dilemma and Development Path of Scientific Research Integrity Management from the Perspective of Data Management

Xia Yikun Qian Jinlin

**Abstract:** [Purpose/significance] Scientific research integrity is the cornerstone of scientific and technological innovation. With the popularization of data-driven research paradigm, the proportion of scientific research misconduct such as data falsification and tampering has increased significantly, highlighting the close internal relationship between scientific research integrity management and scientific research data management.[Method/process] This paper comprehensively uses system analysis method, case analysis method and comparative analysis method to analyze the interactive strengthening effect and gradual accumulation effect of the linkage development of scientific research integrity management and scientific research data management, and then discusses the current predicament of scientific research integrity management from the six elements of subject, object, rule, technology, community and division of labor and their interactive relations under the guidance of activity theory.[Result/conclusion] This paper builds an explanatory framework for scientific research integrity management from the perspective of data management, and puts forward a joint development path of integrity management and data management, including promoting the overall layout of scientific integrity management with complete data system, promoting the system construction of integrity management with high data quality, and deepening the internal driving force of scientific integrity management with efficient data development.

**Key words:** Scientific research data; Scientific research integrity management; Data management; Scientific research misconduct