

科学数据伦理关键问题研究

胡良霖^{1,2,3} 朱艳华^{1,2} 李坤^{1,2} 胡泊^{1,2} 王璐^{1,2} 高瑜蔚^{1,2,4} 李国庆^{5,6}

(1. 中国科学院计算机网络信息中心, 北京 100083; 2. 国家基础学科公共科学数据中心, 北京 100083; 3. 广州物联网研究院, 广东广州 511458; 4. 首都师范大学, 北京 100048; 5. 中国科学院空天信息创新研究院, 北京 100094; 6. 国家对地观测科学数据中心, 北京 100094)

摘要: 随着数据量的急剧增加, 人工智能等技术快速发展和应用, 数据伦理风险日益凸显。本文调研和分析国外重点研究机构发布的数据伦理原则和数据伦理指导框架以及国内伦理相关法律法规和学者的研究进展。在此基础上提出科学数据伦理研究3个关键问题: 科学数据伦理的内涵与外延、科学数据伦理的基本原则以及科学数据伦理研究指导框架要求。

关键词: 科学数据; 数据伦理; 职业道德; 数据共享

DOI: 10.3772/j.issn.1674-1544.2022.01.002

CSTR: 15994.14.issn.1674-1544.2022.01.002

中图分类号: B82

文献标识码: A

Research on Key Issues of Scientific Data Ethics

HU Lianglin^{1,2,3}, ZHU Yanhua^{1,2}, LI Kun^{1,2}, HU Po^{1,2}, WANG Lu^{1,2}, GAO Yuwei^{1,2,4}, LI Guoqing^{5,6}

(1. Computer Network Information Center, Chinese Academy of Sciences, Beijing 100083; 2. National Basic Science Data Center, Beijing 100083; 3. Guangzhou Institute of Internet of Things, Guangzhou 511458; 4. Capital Normal University, Beijing 100048; 5. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094; 6. National Earth Observation Data Center, Beijing 100094)

Abstract: With the rapid increase of data volume and the development and application of artificial intelligence, data ethical risks have become increasingly prominent. This paper investigates and analyzes the principles of data ethics, the requirements of data ethics guidance framework issued by foreign research institutions, and the release of laws and regulations related to domestic ethics and the research progress of scholars. It puts forward three key issues of scientific data ethics research: the connotation and extension of scientific data ethics, basic principles and research ethical guidance framework requirements.

Keywords: scientific data, data ethics, professional ethics, data sharing service

作者简介: 胡良霖 (1973—), 男, 中国科学院计算机网络信息中心研究员, 主要研究方向为科学大数据管理、开放共享与应用服务 (通信作者); 朱艳华 (1982—), 女, 中国科学院计算机网络信息中心高级工程师, 主要研究方向为科学数据管理与应用服务; 李坤 (1987—), 女, 中国科学院计算机网络信息中心工程师, 主要研究方向为科学数据服务与应用; 胡泊 (1992—), 女, 中国科学院计算机网络信息中心助理工程师, 主要研究方向为科学数据标准与应用服务; 王璐 (1996—), 女, 中国科学院计算机网络信息中心助理工程师, 主要研究方向为科学数据管理与应用服务; 高瑜蔚 (1989—), 女, 中国科学院计算机网络信息中心工程师, 主要研究方向为科学大数据管理与共享、标准制订和知识图谱; 李国庆 (1968—), 男, 中国科学院空天信息创新研究院研究员, 主要研究方向为对地观测数据治理与应用服务。

基金项目: 国家科技基础条件平台中心委托任务课题“国内外科学数据发展动态与科学数据资源发展报告编写”(2021WT03)。

收稿日期: 2021年12月1日。

0 引言

在信息时代,大数据、人工智能等新技术的蓬勃发展促进了科技创新和经济发展。同时,大数据应用的负面效益也日益凸显,如大数据杀熟、隐私数据泄露、APP权限滥用、不良内容推荐等事件和现象层出不穷,人类基于传统生活世界的伦理价值也面临着巨大的挑战。有多大的大数据威力,就有多大的挑战^[1]。然而,目前既缺乏制度和法律对大数据伦理进行底线约束,又未形成完善的大数据使用道德规范体系进行预防和抵御^[2]。近年来,数据伦理已经成为国内外各界关注的热点问题。

伦理(ethics)是指“规则和道理”,其中“规则”(伦)和“道理”(理)是同一概念中的两个方面^[3]。美国《韦氏大辞典》指出,伦理学是一门探讨什么是好、什么是坏以及道德责任义务的学科。作为一种“职业道德”,伦理是个人对行为和机构行为的规范和要求,常常以一套有明确条文的规章制度的形式出现。数据伦理(data ethics)是对数据生产、治理、使用和共享过程中个人和机构需要遵守的社会道德和科学规范,是数据从业人员和机构应该遵从的职业道德准则。科学数据伦理则主要针对研究人员、科研机构以及其他科学数据用户,在收集、管理、使用科学数据进行研究、分析和统计过程中的伦理风险考虑和道德要求。

党的十九届四中全会通过的《中共中央关于坚持和完善中国特色社会主义制度推进国家治理体系和治理能力现代化若干重大问题的决定》,首次将数据增列为生产要素,要求“健全劳动、资本、土地、知识、技术、管理、数据等生产要素由市场评价贡献、按贡献决定报酬的机制”^[4]。郭华东等^[5]认为,在数字经济不断部署,逐渐利用收益惠及全社会时,需要克服数字伦理的障碍,数字经济对数据伦理的要求越来越高。如何降低数据在共享利用过程中的伦理风险,最大程度地发挥数据生产要素的经济价值和社会价值,是推动数据要素市场化配置的重大挑战。

科学数据作为新时代传播速度最快、影响面最宽、开发利用潜力最大的战略性、基础性科技资源,已经成为当前国际竞争的战略高地。在科学数据要素市场化配置过程中,同样需要平衡数据开发利用、开放共享、自由流动、数据权属和数据安全之间的关系,降低共享利用可能引发的伦理风险。因此,研究科学数据伦理具有重要的理论价值与现实意义。本文将对科学数据伦理关键问题进行研究,在分析国内外数据伦理研究现状的基础上,探讨科学数据伦理内涵与外延、科学数据伦理基本原则、科学数据伦理研究指导框架等科学数据伦理关键问题。

1 国外数据伦理研究现状

国外数据伦理研究经历了计算机伦理、信息伦理、网络伦理和大数据伦理、人工智能伦理等发展阶段。人们也从对计算机伦理,如黑客问题、知识产权问题、隐私问题、网络空间建构的伦理问题的关注,发展到对大数据的伦理问题(如数据挖掘和算法歧视等)的关注^[1]。近年来,国外一些研究机构在现有法律法规框架体系下,陆续研究并发布了系列数据伦理原则、指导框架要求(包括制定数据伦理检查清单)。如以英国政府中央数字和数据办公室(Government Digital Service, GDS)发布的《数据伦理框架》、美国联邦总务署(General Services Administration, GSA)发布的《数据伦理框架草案》、英国统计局(UK Statistics Authority, UKSA)发布的《使用地理空间数据进行研究和统计时的伦理考虑》、英国开放数据研究所(Open Data Institute, ODI)开发的数据治理工具“数据伦理画板”等。

1.1 英国政府中央数字和数据办公室发布《数据伦理框架》

2018年6月,英国政府中央数字和数据办公室GDS发布《数据伦理框架》(Data Ethics Framework),旨在引导政府和公共部门合理且负责任地使用数据,帮助数据使用者理解伦理要求,并在项目研究过程中解决数据伦理问题,鼓励有责任的创新^[6]。《数据伦理框架》主要的使用

对象是公共部门直接或间接使用数据的人，如数据使用者（统计员、分析师和数据科学家）、政策制定者、政策执行人员和利用数据进行辅助决策的人。该框架认为，在规划、实施和评估项目的过程中，特别是在数据采集、存储、分析或共享发生变化的情况下，研究团队成员应该共同理解并完成数据伦理框架要求，定期回顾框架核心环节。《数据伦理框架》主要是在各关键环节设置问题，逐步引导数据使用者在项目研究中进行数据伦理问题的思考和反思，并通过记录答案完成自我评估的过程。

GDS发布的《数据伦理框架》分为总体原则和具体举措。其中，总体原则适用于项目研究全过程，是项目需要遵循的基本原则，包括透明度、问责制和公平性。原则内容及打分说明详见表1。具体举措则指导研究人员考虑具体数据伦理问题。如定义并充分理解公共利益和用户需求、邀请更多领域专家参与项目过程研讨、在数据使用过程中遵守相关法律和准则、及时审查数据的质量和局限性、制定能够持续评估数据合理使用的计划等。

1.2 美国联邦总务署发布《数据伦理框架》

2019年6月，美国管理和预算办公室（Office of Management and Budget, OMB）发布“2019—2020联邦数据战略行动计划”草案。该草案的核心目标是“将数据作为战略资源（Leveraging Data as a Strategic Asset）”，标志着美国对数据的关注由技术转向资产。同年12月，

美国政府发布《联邦数据战略2020年行动计划》，对行动计划草案进行调整和补充，最终确定各政府机构2020年需采取的20项关键行动，其中第14项行动是开发数据伦理框架。2020年9月，美国联邦总务署GSA发布《数据伦理框架草案》（draft of the Data Ethics Framework）^[7]。其目的是帮助联邦机构系统地识别和评估其在数据获取、管理和使用过程中潜在利益和风险，指导从事收集、管理和使用数据的联邦雇员做出符合道德伦理的决策，为履行其机构使命提供支撑。GSA《数据伦理框架草案》提出7条数据伦理基本原则（表2）。

该框架草案提供了一套可以指导机构开展数据活动的伦理指导原则，有助于数据使用人在决策时能够及时发现数据使用局限、数据利用偏差，以及帮助处理数据使用中涉及的道德问题。任何机构（包括政府部门）在使用数据时，应该冷静思考为社会和公众生活带来的冲击，应充分考虑数据使用不当可能引发的伦理冲突和不良后果，并提出合理且具有前瞻性的数据伦理原则，反向推动法规和技术应用机制的完善。

1.3 英国统计局提出使用地理空间数据的伦理考虑

2021年5月18日，英国统计局（UK Statistics Authority）发布了《使用地理空间数据进行研究和统计时的伦理考虑》（Ethical considerations in the use of geospatial data for research and statistics），探讨使用地理空间数据进行研究、分

表1 GDS《数据伦理框架》总体原则

序号	原则名称	原则内容	打分说明
1	透明度	透明度是指以完整、公开、可理解、易获取和免费的格式发布有关项目的信息，方便研究过程和数据利用公开接受检查	对项目透明度从0到5打分。0分表示项目信息、方法和结果是不开放的，5分表示项目信息、方法和结果对公众是开放的
2	问责制	问责制是指项目具备有效的监督机制。公众或其代表能够对政府及其官员所做的决定和举措进行有效监督，保证政府采取的措施能够实现其既定目标，并满足公共利益	对项目问责制从0到5打分。0分表示项目的监督、管理或同行评审机制尚未建立，5分表示项目已经建立长期的公众监督机制
3	公平性	公平性应考虑消除可能影响模型结果的因素以及对个人和社会群体可能的歧视性影响。确保研究结果尊重个人尊严，是公正的、非歧视性的，符合人权等公共利益诉求	对项目公平性从0到5打分。0分代表该项目存在重大风险，会对公众或某些群体造成伤害或歧视性影响，5分代表项目对人权进行了充分考虑，能够促进公正公平结果，其不利影响可以忽略不计

表2 GSA《数据伦理框架草案》基本原则

序号	原则名称	原则内容
1	遵守法律法规和行业道德准则	数据使用人员应遵守相关的法律规定和行业自律要求
2	诚实并正直行事	所有联邦领导和雇员都应在数据处理过程秉持诚实和正直, 不应该执行或容忍数据使用时的不道德行为
3	实行问责制	问责制要求任何收集、管理或使用数据的人都应了解数据利益相关方并对其负责
4	保持透明度	保持透明度需要明确数据相关的所有活动, 建立与利益相关方沟通的渠道, 并及时更新数据使用的过程和结果
5	跟踪并掌握数据科学领域新技术	要保证所有雇员能够掌握新技术并具备数据处理能力; 及时跟踪数据科学领域发展进展, 特别要关注数据收集、管理和使用的新方法、新技术, 如人工智能、机器学习、神经网络、物联网、区块链等
6	尊重隐私和遵循数据使用限制	使用数据时应尊重人的权利和自由, 保护个人隐私和遵循数据使用限制, 同时遵守相关法律规定
7	尊重公众、个人和社区	所有联邦机构领导和雇员在从事数据活动时, 应从“公共利益”角度出发考虑数据使用的潜在影响

析和统计时的伦理考虑^[8]。在此之前, 为了帮助研究人员和统计人员解决研究项目潜在的伦理问题, 英国统计局制定了一系列数据伦理原则(表3)。

地理空间数据, 也称为“位置数据”或“空间数据”, 是表示人或物体相对于地球表面的位置的数据。使用地理空间数据可以有效服务于公共决策, 如规划当地交通和住房、监测天气并规划防洪等。尽管地理数据带来了巨大利用价值, 但在使用过程中也存在伦理风险和隐私问题。地理空间数据通常用于定位个人地址或企业位置, 属于一种具有特殊性和私密性的数据。利用地理空间数据首先要遵守数据伦理原则。此外, 还要考虑地理空间数据的保密性要求和披露风险、确保地理空间数据的包容性、避免基于地理空间位置引发的偏见和歧视、地理空间数据呈现方式应考虑客观公正等内容。该伦理框架还为使用地理空间数据的研究人员提供了一份详细的数据伦理

检查清单(表4)。

1.4 数据伦理画板

2021年5月, 英国开放数据研究所(Open Data Institute, ODI)针对科研人员数据伦理应用需求, 开发了数据治理工具“数据伦理画板”(Data Ethics Canvas), 以引导个人以合理的方式收集、使用和分享数据^[9]。数据伦理画板可以指导使用者在项目整个管理周期, 对所收集、使用和分享数据行为不断进行思考, 以消除或减轻项目对人们或社会可能带来的潜在危害和风险。ODI画板涉及数据来源、数据版权、数据影响因素、伦理及法律背景等15方面的问题, 协助使用者对数据伦理问题的持续思考和反思(表5)。

同时, ODI给出了数据伦理画板工具详细的使用指南: ①在使用工具前需要考虑利益相关方、邀请专家参与、举办工具培训会。②在使用画板过程中, 可以从画板的任意问题开始, 并与

表3 UK Statistics Authority 数据伦理基本原则

序号	原则名称	原则内容
1	公益事业	使用数据为用户带来明显的益处, 并服务于公共利益
2	保密和数据安全	数据主体的身份(个人或组织)受到保护, 保障信息保密和安全, 并适当考虑知情同意
3	方法和质量	考虑新技术的风险和限制并有足够的监督环节, 使采用的方法符合诚信和质量标准
4	合法合规	使用数据和采用方法要符合法律要求, 如数据保护立法、1998年人权法、2007年统计和注册服务法以及普通法中的保密义务
5	公众参与度	根据所使用数据和研究带来的收益, 考虑公众意见
6	透明度	数据的访问、使用和共享是透明的, 同时与公众的沟通要清晰和方便

表 4 UK Statistics Authority 地理空间数据伦理检查表

序号	检查项	检查内容
1	整体考虑	从项目全局出发，是否仔细思考将要进行的工作可能出现的伦理问题？
2	不伤害	这项工作可能对世界产生哪些积极影响？
3	公益	是否考虑并清楚记录使用地理空间数据给公众带来的价值？
4	方法和质量	是否考虑研究所选择的数据、方法和技术的局限性，将如何进行记录？
5	透明度	是否考虑数据收集、使用、存储和共享的透明度，以及将如何发布研究结果和方法？
6	法律合规	是否考虑与使用数据有关的法律法规？
7	公众意见参与度	是否考虑地理空间数据的特定用途以及潜在公众意见，如何与受影响的群体进行互动？
8	保密和安全	是否考虑维护数据机密性以及安全性的措施？
9	地理位置的选择	所选择的地理位置是否能够避免泄露机密信息，并通过差异化处理降低数据披露风险？
10	按地点披露	是否考虑位置披露对团体或个人造成负面影响的可能性？
11	确保包容	在进行数据分析时，是否考虑所选择数据的完整性或代表性，并在结果中进行明确记录？
12	避免结果偏差	是否考虑研究过程中因为数据选择对结果造成偏差的可能性？
13	意外后果	将采取哪些措施，避免因数据重用而对个人或群体产生不利影响？
14	仔细检查结果	是否留意研究出现的意外结果，以及研究中存在未被识别的关系或误解？
15	地图与可视化	地图类型的选择是否适合研究主题？选择的范围、颜色和标识是否反映客观现实？
16	用户感受	是否考虑读者如何阅读这些研究结果，将获得哪些全新的阅读体验？

更多人分享自己的思考与答案，如使用 ODI 提供的在线模板记录想法和答案；与团队成员共同完成画板；关注参与画板成员的动态。^③完成画板之后，可以跟相关人员分享编辑画板的想法和笔记；反思研究工作对个人、团体和社会产生的持续影响；在画板上查看和更新答案；在项目结束时进行回顾和评估，为以后工作提供帮助。

2 国内数据伦理研究进展

相比国外，我国数据伦理研究起步较晚。不过，对临床研究项目和论文进行伦理审查已经成为行业共识。科技伦理是伦理思想在科学研究和技术开发等科技活动中的应用，是科技活动需要遵循的价值理念和行为规范。值得注意的是，2020 年 10 月 21 日，我国成立了国家科技伦理委员会，旨在推动覆盖全面、导向正确、规范有序、统筹协调的科技伦理治理体系建设。在科技伦理方面，近几年陆续出台相关法律法规，其中有部分内容涉及伦理考虑和要求。科学数据作为科研活动的重要产出和后续研究的重要基础，同样需要遵守科研活动的价值理念和行为规范，同时在数据的采集、加工和利用中需要明确新的伦理要求并指导落实。

2.1 法律法规

2016 年 9 月 30 日，《涉及人的生物医学研究伦理审查办法》自 2016 年 12 月 1 日起施行。该办法明确提出涉及人的生物医学研究项目的负责人作为伦理审查申请人，在申请伦理审查时应当向负责项目研究的医疗卫生机构的伦理委员会提交材料包括文献综述、临床前研究和动物实验数据等资料；知情同意书应当包括研究数据和受试者个人资料的保密范围和措施。2018 年 3 月，发布的《科学数据管理办法》虽然没有明确提出科学数据伦理，但是明确要求，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。2021 年 7 月，科技部发布的《关于加强科技伦理治理的指导意见（征求意见稿）》明确了科技伦理的 5 项基本原则，即增进人类福祉、尊重生命权利、坚持公平公正、合理控制风险、保持公开透明，并要求高等学校、科研机构、医疗机构、企业等单位要履行科技伦理管理主体责任，对本单位科技活动的伦理风险主动研判、及时化解。总之，建立健全相应的法律法规和伦理准则，提升网络维护技术手段，加强行为主体的伦理道德教育是应对科学数据伦理问题的

表5 ODI数据伦理画板问题清单

序号	问题项	问题内容
1	数据源	①命名/描述项目的关键数据,包括通过个人收集或第三方访问获取的数据 ②相关数据是否涉及个人隐私或其他敏感信息?
2	数据版权	①您从哪里收集到数据?它是由团队制作的还是个人直接收集? ②收集数据是为了正在完成的项目还是有其他目的?您有使用该数据的权限吗?使用依据是什么?数据拥有哪些永久性版权?
3	影响实验数据 的因素	您在使用数据的过程中,是否存在可能影响实验结果的因素?如数据采集、汇聚、处理、分析过程以及算法演绎方面的偏差;数据的缺失;数据来源和数据质量;影响决策的其他问题,如团队组成等
4	法律背景	您的团队和项目在使用数据时应遵循哪些现行道德规范?您个人使用数据应当遵循哪些法律、政策和法规?具体要求是什么?如法律,人权,数据保护,知识产权和数据库权限,反歧视法,数据共享、政策、条例以及特定部门(如卫生、就业、税收)的道德守则/框架
5	持续执行要求	①您是否经常考虑受项目影响的人的想法?您是如何考虑的? ②需要哪些信息或培训帮助人们理解数据? ③是否有专门的系统、流程或资源应对可能会长期出现的数据问题?
6	数据使用目的	①您在项目中收集和使用数据的主要目的是什么? ②您的主要应用场景是什么?您的商业模式是什么? ③您使用数据开展的工作是否对社会有利?具体对哪些群体有益? ④借助于数据,您当前正在进行的项目是否可以取代其他产品或服务?
7	对社会产生的 积极影响	①这个项目可以对哪些团队和单位产生积极影响?具体哪些积极影响? ②您是如何衡量并传达这些积极影响的?您是如何促进这些积极影响的?
8	对社会产生的 负面影响	①哪些人会受到这个项目的负面影响? ②收集、使用或共享数据的方式是否会带来风险使负面影响重复出现? ③数据的收集和使用是否具有针对性,会对特定人群产生偏见,或者限制特定人群的访问(排他性)? ④数据使用限制性条件及可能产生的风险是如何传达给人们的?
9	最小化负面影 响	①您采取哪些措施能够将数据利用带来的伤害降到最低? ②如何减少使用数据的一些限制条件?您如何保证个人隐私数据以及其他敏感数据的安全? ③您如何评估项目潜在的负面影响?如何记录处理方法? ④消除或减轻负面影响会给您的项目带来什么益处?
10	数据伦理审查 与迭代	①如何发现和处理正在发生的数据伦理问题? ②对于画板的内容,您多久更新一次?什么情况下进行更新?
11	与公众互动	①他人如何就您的项目与您进行互动? ②他人如何修改数据以及申诉或要求更改产品/服务?具体应怎样做? ③申诉机制是否合理?是否能够被人们充分理解?
12	交流使用数据 的目的	①数据的利益相关方是否理解您使用该数据的目的? ②您是否已清楚地向他人传达您使用该数据的目的? ③您如何确保相关弱势群体理解您的数据使用目的?
13	保持开放和透 明度	①您的项目通过什么方式开放共享?您是否公布项目采用的方法、元数据、数据集、代码以及标准? ②您将如何与同行交流以获得反馈? ③您会在这张画板上公开发表您完成画板的方法与答案吗?
14	共享数据	①您会与其他团队共享数据吗?如果是,会与谁共享? ②您打算公开发布该数据吗?在什么条件下发布数据?
15	确定行动及对 应负责人	①关于数据伦理,在项目开始之前您会先采取什么行动?执行顺序是什么? ②哪些人将为这些行为负责?哪些人必须参与此项行动?

有效途径之一^[10]。

2.2 学术探讨

与国外较早的关注和丰富的研究成果相比，国内研究相对滞后且研究成果较少，没有形成较为明显的研究阶段^[11]。笔者在同方知网学术文献总库中选择了期刊、博士论文、硕士论文、会议论文和报纸等五类文献资源，以数据伦理为检索词，在篇名字段中进行精确检索，检索时间限定为近5年（2016年1月1日至2020年12月31日）。最终共获得以数据伦理为篇名的文献35篇。①从论文发表年度看，数据伦理论文逐年增多。其中，2020年有12篇、2019年有11篇、2018年有6篇、2017年有6篇。②从主题聚类看，数据伦理论文研究主要集中在大数据（16篇）、伦理问题（9篇）、数据伦理（5篇）、大数据伦理（4篇）、伦理失范（3篇）等主题内容。③从学科分布来看，数据伦理论文主要集中在伦理学（14篇）、计算机软件及计算机应用（8篇）、新闻与传媒（4篇）、民商法（3篇）、信息与邮政经济（3篇）、医学教育与医学边缘学科（3篇）等学科领域。

深入分析相关文本发现，研究者关于“数据伦理”的讨论主要集中在大数据面临的主要伦理挑战、数据伦理失范产生的原因分析、数据全生命周期关键环节面临的伦理问题、数据伦理涉及的责任主体以及学科领域（特别是医疗健康领域）利用数据带来的伦理问题等方面。其中，医疗健康领域由于涉及较多的个人数据，特别是个人隐私和敏感信息，面临更多的数据伦理风险。如医患关系信任危机、数据泄露风险、数据被过度商业使用等。随着医疗技术的提升，医疗数据获取更加方便，患者有可能成为医生眼里由检查数据构成的数字人体，在一定程度上忽略了患者的心理感受，医学人文关怀无法体现。同时，医疗信息商业化更是令人担忧，有调查显示医疗行业的数据泄漏问题要远高于其他行业^[12]，并以《医学科学数据共享与使用的伦理要求和管理规范》为题进行了深入探讨^[13-24]。此外，针对大数据伦理治理责任主体，有研究者认为责任主体涉

及个体、行业、国家乃至国际多个层面。其中，个体主体要培养良好的数据管理行为，行业企业大数据伦理治理离不开行业的规范和自律，而数据权属问题、公众隐私权、遗忘权和反歧视等问题则需要从国家层面通过法律法规予以确立和保障^[1]。

3 科学数据伦理关键问题分析

科学数据伦理是一个相对复杂的问题，涉及数据技术、法律法规以及个人信息素养等。所有相关人员在科学数据工作的过程中，都应积极开展数据伦理思考，降低数据利用带来的伦理风险，以实现公共利益的最大化。特别是对科学数据伦理的内涵与外延、科学数据伦理的基本原则以及科学数据伦理应用指导框架等关键问题值得进行研究探讨。

3.1 科学数据伦理内涵与外延

关于数据伦理的定义目前尚未形成统一的表述。许多论文和出版物对数据伦理的内涵进行了不同语境下的解释和讨论。美国联邦总务署GSA认为，数据伦理是指在收集、管理或使用数据时，为实现保护公民自由、最大程度地降低个人和社会的数据使用风险以及实现公共利益最大化等目的，进行适当判断和问责的依据。英国开放数据研究所ODI认为，数据伦理是道德的一个分支，用于评估人们在数据收集、共享和使用过程中可能对人和社产生影响的数据实践。还有学者探讨了科学技术、伦理和法律的边界，认为科学技术是对事物做一个事实的判断，解决的是这件事情能不能做；伦理是对这个事物进行价值判断，解决的是这件事情应该做还是不应该做；而法律是对社会安全的一个判断，解决的是这件事情准不准做^[25]。

除了界定科学数据伦理的定义和内涵外，值得探究的问题还包括科学数据伦理面临风险与挑战（如从个人到弱势群体，再到涉及所有人的公共利益等）；引发这些风险的内在动机（如技术引发的算法偏见、商业利益驱动的数据滥用等）；伦理风险带来的不良后果与影响（如隐私暴露风

险对个人带来的精神压力和伤害)；科学数据伦理涉及的主要责任主体(如政府管理部门、组织机构、数据使用者等)。

3.2 科学数据伦理基本原则

科学数据伦理基本原则是伦理得以实施的基本准则和认知共识。关于数据伦理基本原则,如前文所述,英国政府中央数字和数据办公室发布的《数据伦理框架》提出了3项原则(透明度、问责制、公平性),美国联邦总务署发布的《数据伦理框架草案》提出7项原则(遵守法律法规和行业道德准则、诚实并正直行事、实行问责制、保持透明度、跟踪并掌握数据科学领域新技术、尊重隐私和遵循数据使用限制、尊重个人公众和社区),英国统计局提出数据伦理6项原则(公益事业、保密和数据安全、方法和质量、合法合规、公众参与度和透明度)。虽然在表述方式和排列顺序存在不同,但有些数据伦理原则已经达成共识,如透明度、公平性等。此外,《关于加强科技伦理治理的指导意见(征求意见稿)》明确提出在制订科学数据伦理基本原则时,需要考虑科技伦理的5项原则(增进人类福祉、尊重生命权利、坚持公平公正、合理控制风险、保持公开透明)。

本文也提出了科学数据伦理的5项基本原则,即科技向善(Social Good)、以人为本(Human-Oriented)、透明(Transparency)、公开(Openness)、公正(Fair)。其中,科技向善要求在使用科学数据的过程中,最大限度地减少对人的生命安全和心理健康等可能造成的潜在伤害或威胁。以人为本原则要强调利用科学数据要追求为社会和人类福祉做出贡献,要尊重人格尊严、个人隐私和知识产权等权益。透明原则要求客观记录数据来源、数据用途等必要信息,保证数据内容和利用程序的透明度,以及充分考虑社会公众的合理参与。公开原则要秉承科学数据完整客观、来源准确、数据引用真实。公正原则是在使用科学数据的过程中做到公平无歧视,充分考虑不同的社会群体,特别是弱势群体的权益,尽量缩小数据鸿沟。

3.3 科学数据伦理研究指导框架

科学数据伦理指导框架是科研人员在利用科学数据开展日常工作,进行伦理决策的核心参考和要求。目前,国际上比较通用的做法是针对项目研究中可能遇到的数据伦理问题设计问题清单,并通过问卷整理和反思在数据过程中的伦理风险。如英国开放数据研究所ODI在数据伦理画板中设计了16个伦理问题清单,英国统计局UKSA发布的地理空间数据伦理问题清单包括17个问题。这些问题涉及研究人员使用科学数据开展科研项目的各个阶段(如项目开始之前、数据使用过程、研究结果发布等)。国内有学者整理了数据伦理参考准则通用分析框架,考察科研人员在进行数据开放共享过程中,利己主义因素、功利主义因素、利他主义因素及实用主义因素对数据开放共享带来的影响^[26]。

科学数据伦理框架涉及了较多的利益相关方,不同学科之间的伦理共识也存在差异。本文仅提出科学数据伦理研究指导框架五要素模型,包括研究目的、研究对象、数据选择、算法工具和应用场景。每个过程都有可能涉及数据伦理问题和伦理风险,数据使用者需要不断平衡法律法规和职业道德、个人利益和社会利益之间的关系(图1)。

- ①针对研究目的和预期目标,要综合考虑科学数据伦理基本原则以及各研究阶段可能面临的伦理风险。
- ②针对研究对象,要充分考虑个人、弱势群体以及涉及所有人的公共利益诉求。
- ③在选择数据时,要禀赋责任的严谨态度,充分考虑数据的真实性、可靠性、样本数据的代表性,并严格遵守相关平台标准规范,对数据进行描述和说明。
- ④在选择算法和工具时,要警惕算法偏见和歧视。一类来自历史数据或新采集数据集的偏见,经过算法再一次被放大。另一类来自算法工程师的个体偏见。如果算法工程师自身抱有对性别和种族的歧视,就有可能把个人的偏见和歧视渗透到算法中^[1]。
- ⑤在数据应用过程中,要充分考虑数据被误用和被滥用的风险,还要遵守文献引用习惯,进行规范的数据引用,承认他人的科学贡献,表达对数据创造者的认可和

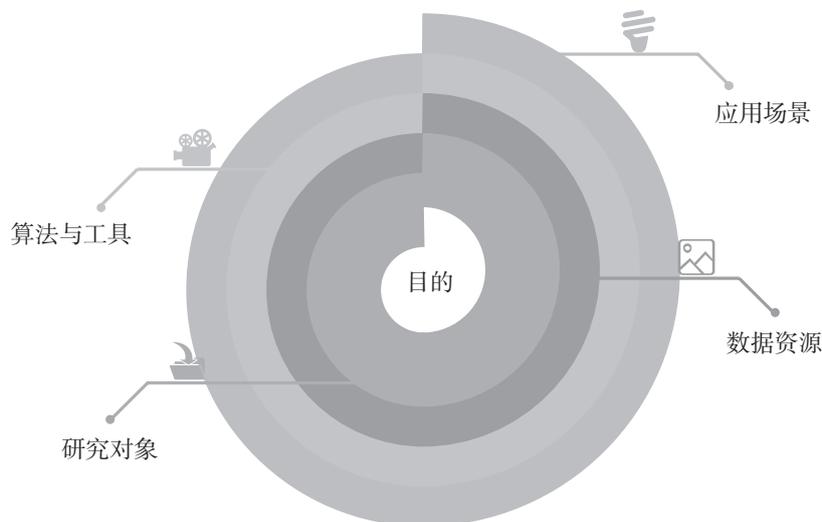


图1 科学数据伦理研究指导框架五要素模型

尊重。

4 结语

科学数据伦理是对使用数据的人或机构提出的职业道德要求。本文在调研和分析国内外数据伦理研究现状的基础上，提出了科学数据伦理面临的3个核心问题，即科学数据伦理内涵与外延、科学数据伦理基本原则和科学数据伦理研究指导框架。同时，提出科学数据伦理的5项基本原则以及科学数据伦理研究指导框架五要素模型。科学数据伦理问题涉及较多的利益相关方，同时不同学科领域的伦理共识也不尽相同，需要结合具体场景进行进一步的研究和分析。

参考文献

- [1] 中国科学院. 2020 高技术发展报告 [M]. 北京: 科学出版社, 2021.
- [2] 田维琳. 大数据伦理失范问题的成因与防范研究 [J]. 思想教育研究, 2018(8): 107-111.
- [3] 王冬桦. 为伦理与道德的概念及其关系正本清源 [J]. 首都师范大学学报(社会科学版), 2011(2): 119-124.
- [4] 中国共产党十九届四中全会新闻发布会 [EB/OL]. (2019-06-19)[2021-09-13]. http://www.gov.cn/xinwen/2019-11/01/content_5447495.htm.
- [5] GUO H D, MICHAEL F G, ALESSAND R O A. Manual of digital earth [M]. Singapore: Springer, 2019.
- [6] Data Ethics Framework [EB/OL]. (2019-06-19)[2021-

06-05]. <https://www.gov.uk/government/publications/data-ethics-framework>.

- [7] draft of the Data Ethics Framework [EB/OL]. (2019-06-19)[2021-05-24]. <http://www.meritalk.com/articles/gsa-drafts-fed-data-ethics-framework/>
- [8] Ethical considerations in the use of geospatial data for research and statistics [EB/OL]. (2021-06-19)[2021-07-23]. <https://uksa.statisticsauthority.gov.uk/publication/ethical-considerations-in-the-use-of-geospatial-data-for-research-and-statistics/pages/1/>.
- [9] Data Ethics Canvas [EB/OL]. (2021-06-19)[2021-09-15]. <https://theodi.org/article/the-data-ethics-canvas-2021/>.
- [10] 陆伟华. 大数据时代的信息伦理研究 [J]. 现代情报, 2014, 34(10): 66-69.
- [11] 温亮明, 张丽丽, 黎建辉. 大数据时代科学数据共享伦理问题研究 [J]. 情报资料工作, 2019, 40(2): 38-44.
- [12] 袁杨, 关键. 健康医疗大数据面临的伦理挑战与应对思考 [J]. 中华医学科研管理杂志, 2019, 32(6): 4.
- [13] 关键. 医学科学数据共享与使用的伦理要求和管理规范(一)前言 [J]. 中国医学伦理学, 2020, 33(2): 143-146.
- [14] 关键. 医学科学数据共享与使用的伦理要求和管理规范(二)隐私变迁与挑战 [J]. 中国医学伦理学, 2020, 33(3): 288-293.
- [15] 关键. 医学科学数据共享与使用的伦理要求和管理规范(三)知情同意履行挑战与原则策略 [J]. 中国医学伦理学, 2020, 33(5): 530-535.
- [16] 关键. 医学科学数据共享与使用的伦理要求和管理规范(四)伦理要求: 基本伦理原则的创新解读 [J]. 中国

- 医学伦理学, 2020, 33(6): 645-649, 683.
- [17] 关键. 医学科学数据共享与使用的伦理要求和管理规范(五)隐私分类分级的初步建议及其依据的确认[J]. 中国医学伦理学, 2020, 33(8): 915-920.
- [18] 关键. 医学科学数据共享与使用的伦理要求和管理规范(六)机构伦理管理的专家建议[J]. 中国医学伦理学, 2020, 33(9): 1031-1034, 1045.
- [19] 关键. 医学科学数据共享与使用的伦理要求和管理规范(七)伦理审核指南专家共识[J]. 中国医学伦理学, 2020, 33(10): 1159-1166.
- [20] 关键. 医学科学数据共享与使用的伦理要求和管理规范(八)审核委员会职责和高效的审核机制探讨[J]. 中国医学伦理学, 2020, 33(11): 1306-1310.
- [21] 关键. 医学科学数据共享与使用的伦理要求和管理规范(九)医学科学大数据及其属性[J]. 中国医学伦理学, 2020, 33(12): 1427-1432.
- [22] 关键. 医学科学数据共享与使用的伦理要求和管理规范(十)大数据产权认定解决方案的建议[J]. 中国医学伦理学, 2021, 34(1): 22-26.
- [23] 关键. 医学科学数据共享与使用的伦理要求和管理规范(十一)重大传染病数据共享应用挑战和潜在审核方案[J]. 中国医学伦理学, 2021, 34(2): 131-136.
- [24] 关键. 医学科学数据共享与使用的伦理要求和管理规范(十二)通用伦理准则要点建议及其注释[J]. 中国医学伦理学, 2021, 34(3): 280-287.
- [25] 临床科研项目都需要伦理审查吗? [EB/OL]. (2019-06-19)[2021-10-15]. <https://mp.weixin.qq.com/s/SyOFheQ0TDzOzLN3bZNHQw>.
- [26] 刘金亚, 顾立平, 张潇月, 等. 开放科研数据环境下科研人员的数据伦理框架研究[J]. 情报理论与实践, 2020, 44(2): 83-89.

(上接第10页)

- [2] 和海霞, 陈伟涛, 吴玮, 等. 灾害事件和灾害过程驱动下Charter数据获取机制研究[J]. 国土资源遥感, 2014, 26(2): 190-195.
- [3] United Nations Platform for Space-based Information for Disaster Management and Emergency Response(UN-SPIDER)[EB/OL].(2010-11-10)[2021-10-30].<http://www.unoosa.org/oosa/en/our-work/un-spider/index.html>.
- [4] 赖文, 康馨尹. UN-SPIDER联合国外空司灾害管理与应急响应天基信息平台[J]. 中国减灾, 2015(17): 31.
- [5] 李素菊. 空间减灾国际合作机制(三)联合国灾害管理与应急响应天基信息平台机制介绍[J]. 中国减灾, 2018(13): 56-57.
- [6] The Operational Satellite Applications Programmer of the United Nations Institute for Training and Research(UNITAR/UNOSAT)[EB/OL]. [2021-10-30]. <https://unitar.org/maps/countries>.
- [7] International Charter Space and Major Disasters (Charter)[EB/OL].(2000-11-01)[2021-10-30]. <https://disasterscharter.org/web/guest/home>.
- [8] 王志刚, 支持国际灾害救援, 贡献中国航天力量: “空间与重大灾害国际宪章”实施20周年[J]. 卫星应用, 2020(11): 14-20.
- [9] 郭朝辉, 邓佳楠. 履行Charter ECO职责, 持续支援国际灾害应急响应[J]. 卫星应用, 2021(1): 44-49.
- [10] Copernicus Emergency Management Service(CEMS)[EB/OL].[2021-10-30].<https://emergency.copernicus.eu/>.
- [11] Sentinel Asia[EB/OL].[2021-10-30].<https://sentinel-asia.org/>.
- [12] 赵展一, 黄金霞. 开放科学基础设施的信息资源建设模式分析[J]. 图书馆建设, 2021(3): 46-55.
- [13] GEO Data Sharing Principles[EB/OL].(2009-11-17)[2021-10-30].<http://www.earthobservations.org/dswg.php>.
- [14] WILKINSON M D, DUMONTIER M, AALBERS-BERG I J, et al.The FAIR guiding principles for scientific data management and stewardship[J].Scientific data, 2016, 3(1): 167-172.
- [15] GEO Data Management Principles[EB/OL].(2015-04-28)[2021-10-30].<http://www.earthobservations.org/dswg.php>.