研究论文
DOI：10.13998/j.cnki.issn1002-1248.23-0158

可重复性危机背景下的医学数据素养教育体系研究

孔祥辉，孙　璞*
（锦州医科大学，锦州　121017）

摘　要：［目的／意义］目前生物医学研究正面临可重复性危机，提升科研人员的数据素养成为了化解这场危机的关键所在。本研究揭示可重复危机与数据素养及其教育的概念联系，拓展教育新内涵、并分析构建教育运行体系和保障因素，为中国开展相关教育提供参考。［方法／过程］提出可重复性数据素养教育新内涵，在充分总结国外教育实践成果的基础上，从目标与内容、主体与客体、教学方式、实施策略、评估与评价角度，分析构建可重复性数据素养教育体系，并分析体系运行的保障因素。［结果／结论］可重复性数据素养教育以提升研究可重复性为价值目标，内容划分为数据意识、数据技能、数据伦理 3 个维度；医学图书馆作为教育核心主体，应面向广大受众群体建立多维立体的培养方式，坚持主动学习的课堂实施策略与多元化评估与评价手段来开展教育，并通过加强队伍建设、夯实资源基础、推动教学交流、完善制度建设以保障教育的顺利实施。

关键词：数据素养教育；数据管理；可重复性；医学信息学；医学图书馆
中图分类号：G258.2　　**文献标识码：**A　　**文章编号：**1002-1248（2023）04-0070-09

引用本文：孔祥辉, 孙璞. 可重复性危机背景下的医学数据素养教育体系研究[J]. 农业图书情报学报, 2022, 35(4)：70-78.

1　引　言

可重复性（Reproducibility）是指利用与原研究相同的数据、代码、方法、步骤和条件获得相同结果的能力[1]。可重复性是现代科学的基石之一，科研的创新性必须建立在稳定的、可以被反复验证的结果基础上才具有意义。然而，现代生物医学科研领域正饱受可重复危机（Reproducibility Crisis）的困扰，存在大量虚假的研究成果无法重现[2]，由此导致了学科知识积累缓慢，临床试验疗法失败、药物研发受阻、科研经费浪费等诸多问题的产生。

目前有关可重复性危机的研究成果集中于分析危机所带来的影响、追溯成因并从不同角度探寻解决对策的研究。如 MULLANE[3]认为可重复危机严重破坏了整个生物医学研究生态系统的可信度和可持续性，而研究者思维方式改变将有助于问题的缓解。BEGLEY[4]认为可重复性危机是一个涉及面广、与多方利益攸关的复杂问题。没有一方可以单独负责，也没有单一且一劳永逸的解决方案。PAPAMOKOS[5]、VOELKL[6]等深入剖析对生物医学研究可重复性构成威胁的因素，主要包括缺乏科学严谨性、统计能力低、生物材料复杂性、分析灵活性和欺诈。SAMSA[7]从数据管理角度提出解决方案，强调应提前制定数据分析计划、清晰

收稿日期：2023-03-14
基金项目：2022 年度辽宁省社会科学规划基金青年项目“可重复性危机视域下医学院校图书馆数据素养教育研究”（L22CTQ004）
作者简介：孔祥辉（1987-），硕士，锦州医科大学图书馆，馆员，研究方向为信息服务
*通信作者：孙璞（1987-），硕士，锦州医科大学药学院，组织员，研究方向为教育管理。Email：573223798@qq.com

的数据管理和分析协议、详细的实验方案、实施积极的实验室管理实践，用来支持“严谨 + 透明度 = 可重复性”的文化构建。

现代生物医学研究已进入数据密集型科研范式，数据成为支撑科学结论 / 结果的核心要素，因此必须认识到可重复性危机的实质是关于数据行为能力与规范的危机。努力改善科研人员的数据实践行为，提高数据质量才是化解危机的关键。而数据素养教育作为有目标、有组织、有系统、有计划、有评价的教育活动，通过提供系统化的教育内容指导科研人员开展正确、专业、高效、合理、合规的科研数据实践活动，使之提高科研能力以适应现代生物医学研究方式，最大程度提高数据产出和质量，成为危机治理的有效手段。

从数据素养教育角度切入可重复性危机化解的研究文献较少，ROCHE[8]、SAMUEL[9]等强调了危机环境下研究数据管理培训和课程建设对于改善数据共享实践，提高数据的透明度和可重用性的重要性。部分成果[10-12]也介绍了教学实施案例和项目，但缺乏从整体视野审视危机时代所赋予数据素养教育的新使命、并构建全新的教育体系，形成系统化方案指导实践。本文力图克服现有研究不足，揭示可重复危机与数据素养及其教育的概念联系，拓展教育新内涵、并分析构建教育运行体系和保障因素，为推动当前数据素养教育转型和创新提供参考。

2　理论基础

2.1　可重复危机与数据素养

数据素养是获取、解释、评估、管理、处理和合理利用数据的综合能力。内在包括 3 个维度[13]：①数据意识。对科学数据的认知程度。包括数据敏感性、价值认知、数据共享与协作意识。②数据技能。涵盖数据生命周期的数据管理能力，包括收集、分析、处理、出版、组织、保存、评价与再利用等阶段的知识和技能。③数据伦理。数据实践中所遵循的规范和道德准则。包括数据规范、数据隐私、数据保护等内容。

可重复危机是整体领域研究状态的宏观描述，表现于众多微观个体研究成果的不可重复性（Irreproducibility），是其本身数据、方法、过程、环境、结果 5 个要素共同作用下的结果。如果任何一种要素存在潜在风险和问题，就无法确保外界拥有与原研究相一致的状态，在重复实施下获得相同结果。而引发这些风险或问题又是源于科研人员存在的数据实践行为问题（表 1）。

表 1　研究不可重复因素及其成因分析

Table 1　The Influencing factors and cause analysis of Irreproducible Research

因素	潜在风险&问题	数据行为原因
数据	存在错误缺陷	缺乏质量控制（Insufficient Quality Control）[14]：未能识别和解决数据收集和输入中潜在的错误
	不可识别	描述不充分（Inadequate Description）[15]：未能充分利用元数据进行注释，为原始数据和方法提供上下文和来源信息
	无法完整共享	数据监管不佳（Poor Data Curating）[16]：无法对研究过程中的广泛、复杂的数据资产（包括数据集、代码、模型、文章、预印本、协议）形成有效监管
方法	无效或错误	*p* 值误用（Misuse of P-Values）[17]：将 *p* 值作为判断假说真伪和结果重要性的唯一依据
过程	缺乏完整清晰	出版透明度不足（Insufficient Transparency of Reporting）[16]：无法公开、清晰和全面地报告和传播数据分析结果。研究报告中有关研究方法和数据分析的关键信息缺失、不一致、不完整或具有误导性
	不确定性增加	分析灵活性（Analysis Flexibility）[18]：指在数据预处理和统计分析期间做出的大量选择，影响分析结果或解释
环境	差异性巨大	缺乏变量控制（Lack of Variability Control）[19]：不能准确识别、控制、记录实验进程中涉及的变量条件，忽视包括盲法、随机化、复制、样本量计算和性别等关键实验元素所带来的差异性影响
结果	不可靠	缺乏复制研究（Lack of Reproducible Research）[20]：缺乏对于自身成果进行再次检验
	不客观	发表偏见（Publication Bias）[21]：过度追求发表积极的、具有统计学意义的数据结果，放弃负面发现。甚至扭曲成果
	真实性存疑	可疑的研究实践（Questionable Research Practices）[17]：是指为实现预期结论，在数据设计、分析或报告中采取的可能会产生偏见的行动，包括选择性报告、*P* 值操纵、已知结果假设、删除异常值、伪造数据等

DOI：10.13998/j.cnki.issn1002-1248.23-0158

数据素养是科研人员自身稳定、内在的状态，数据实践行为则是数据素养的外显表达。数据素养的不足容易导致科研人员各类数据实践的行为问题，造成原研究的各要素存在诸多潜在风险和问题，最终导致不可重复性（图 1）。

2.2 可重复危机背景下数据素养教育新内涵

医学数据素养教育是结合生物医学领域学科特点，为培养该领域科研人员数据意识，提高数据技能和道德水平而开展的专业化教育。可重复性危机凸显医学数据素养教育的必要性，同时也对教育内容提出了新要求，应将提高研究可重复性作为重要的价值导向，创新数据意识教育，推动科研人员数据观念的切实转变，克服研究偏见，将价值观转向过程而不是结果；建立批判、自省的思维方式和研究文化；以开放科学数据管理 FAIR 原则为指导，深化基于生命周期的数据技能教育，助力科研人员通过各阶段的最佳实践来降低数据行为风险，切实提高数据可查找、可访问、可互操作和可重用水平，提升方法有效性、过程的完整透明程度、环境条件的可控度、结果的精准可靠程度；强化数据伦理教育，确保科研人员能始终坚守诚信底线，在严格遵循数据伦理规范下实施真实负责任的研究。最终建立可重复性数据素养教育（Reproducibility Data Literacy Education，简称 Re-DLE）体系。

3 Re-DLE 体系构建

3.1 Re-DLE 目标与内容框架

以数据素养维度为依据，Re-DLE 内容可划分为 3 个部分，将各研究要素的可重复性要求细化拆分，结合数据生命周期，提炼出明确的教育目标；借鉴现有文献成果与欧美国家和地区教学实践成果[22-29]，整合形成系统的 Re-DLE 教育内容框架，如表 2 所示。

3.2 Re-DLE 客体与主体

教育客体是教育的接受者。可重复性作为检验科研成果是否正确可靠的黄金标准，是从事生物医学领域科研活动的所有潜在群体的共同价值追求。因此 Re-DLE 的教育客体涵盖了本科生、研究生、博士后、实验员、教职工等不同科研群体人员类型。

教育主体是教育的实施者。Re-DLE 教育内容的广泛性决定了教育主体呈现多元化格局，国外实践表明，生物医学科研院所、基金资助机构、科研管理机构、

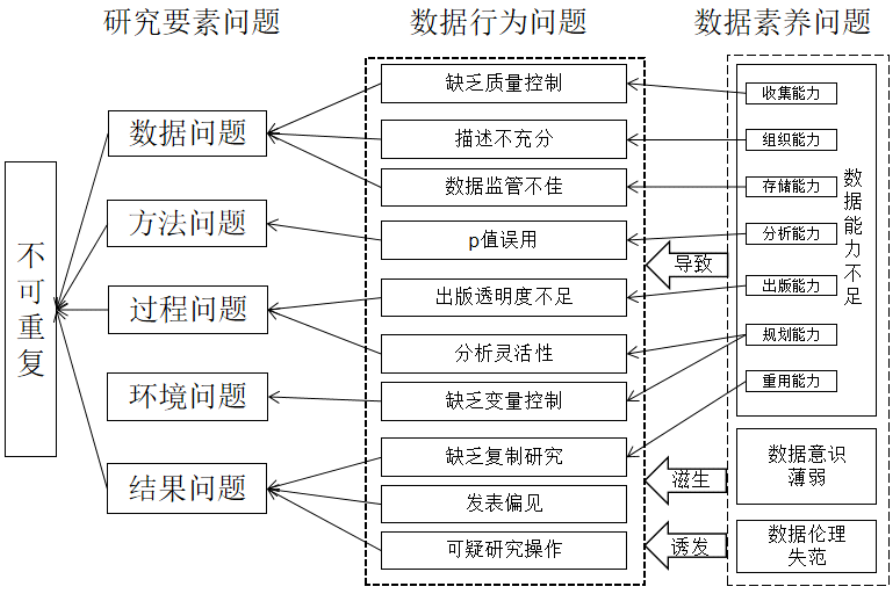

图 1 研究不可重复与数据素养问题关系图

Fig.1 A map of the relationships between data literacy and irreproducible research

表 2　Re-DLE 教育目标与内容框架

Table 2　The objectives and contents of reproducibility data literacy education

维度	教育目标		教育主题模块	具体内容
可重复性数据意识	认知与思维建立		可重复性基础教育	可重复性危机、生物医学可重复性、开放科学、科研数据管理等基础知识
可重复性数据技能	方法	正确有效	统计素养教育	统计方法与规范（效应大小测量、置信区间计算 定量统计方法、统计功效）
	过程	清晰合理	严谨性规划教育	预注册、数据管理计划（DMP）
		公开透明	元数据教育	元数据工具、使用方法、规范、策略
	环境	精准把控	自动化教育	自动化计算、编程、统计与管理工具（版本控制、R、RStudio 等）
	数据	质量控制	标准化教育	数据清洗、标准化处理（动物模型、生物变量控制细胞，抗体等资源认证、样本量处理等）
		完整	开放存储教育	开放数据存储标准、工具、方法与策略（Open Science Framework 等）
		高度共享	开放出版教育	开放数据出版标准、工具、方法与策略（报告指南、预印本、数据论文等）
可重复性数据伦理	结果	可靠	数据复制教育	数据验证与评估方法、工具、项目构建（虚拟机与容器、可重复研究流程）
		真实	学术诚信教育	学术不端、可疑研究操作等问题认识与预防
		规范	数据规范教育	负责任研究的数据伦理规范（数据来源规范、技术规范、开放规范、实践规范等）

学术出版商、医药厂商等一切科研相关利益机构都可以是 Re-DLE 的教育主体。但其中，医学院校图书馆（以下简称医学馆）具有独到的专业优势，一方面通过创建标准元数据、分配标识符、提供文件格式转换和数字化保存等数据管理活动来增强数据集的价值；同时在数据生命周期的各个阶段都能与生物医学研究人员合作，以遵循标准规范和符合伦理要求的方式共享数据并支持研究透明度。另一方面，作为传统医学信息素养教育主要承担者，医学馆可基于数据素养与信息素养内在的概念关联进行教育的传承创新，成为 Re-DLE 的关键教育力量。

3.3　Re-DLE 教学方式

医学馆应作为开展的 Re-DLE 主阵地，建立多维立体的培养方式。坚持做到：①独立与融合统一。既要坚持对 Re-DLE 的主导性，独立承担包括研讨会、培训、学分课程等在内的教学形式。又要重视与其他教育的有效衔接。美国医学馆将 Re-DLE 内容有机纳入到硕博士群体的“负责任研究行为”教育（Responsible Conduct of Research，RCR）或继续医学教育（Graduate Medical Education，GME）体系中，以争取能获得更好的关注及制度保障[30]。如约翰霍普金斯大学将由图书馆和统计专业教师共同主讲的数据伦理与科研诚信课程作为所有研究人员每年必修的继续教育课程[31]。②线上线下结合。既要充分利用医学馆已有教研室、学术报告厅等空间开展各类线下教学，也要利用网络与多媒体技术，将线下教学资源整合建立在线教学平台和资源导航、开设网络公开课、迷你课、微课、网课等形式 突破时空限制，鼓励用户利用网络开展自学、降低教学成本。③层次与个性兼顾。不同类型的科研用户群体所承担分工不同，教育需求也各不相同。要精准匹配与之对应的教育方式和内容。如本科生作为科研储备力量，应嵌入专业课堂开展可重复数据意识教育；研究生是科研活动的参与与协同者，

DOI：10.13998/j.cnki.issn1002-1248.23-0158

应建立系统培训并突出数据技能教育；专门从事科研的职业人员，特别是博士、博士后、实验员等，应深入到实验室或科研环境中提供个性化指导，全面覆盖教学内容并体现足够深度。此外，生物医学包含了众多不同专业，而这些专业的科研数据特性、应用、分析手段的不同，决定了其科研可重复要求不同，必须结合专业背景，合理设置个性化内容方案。

3.4 Re-DLE 实施策略

教学实施策略的多样化可以满足广泛的学习者需求和偏好，并可能对学习过程产生积极影响。医学馆在开展教学过程中，应坚持学生主动学习策略，激发学生学习兴趣，引导学生自主构建相关理论知识与技能。国外目前所采取的有效经验有：①任务驱动教学。根据教育内容所涵盖的主题设计对应的实践任务，例如可设计复制任务，要求学生对经过同行评审的论文出版物进行评估，在掌握必要的技术参数、工具、方法步骤进行成果复制，在这一过程中逐步体会提升可重复性的优势、熟悉流程构建，并识别潜在障碍，进而实现数据复制教育[32]。②小组团队协作。以 3~5 人的小组形式，共同完成各类话题讨论和任务，锻炼学生分析、推理、评估、沟通和团队合作能力。③开展游戏或竞争比赛。设计游戏将可重复知识和技能概念融入其中，如荷兰莱顿大学图书馆举办 ReproHack（可重复性黑客马拉松），鼓励同学使用可访问的代码和数据重现已发表的论文，将重现程度和效果作为竞争依据，以此激活学生兴趣和创造力[33]。④主题扩展阅读。围绕主题扩展阅读资源，例如，榜样人物、个人故事、期刊社论、新闻专栏、公共演讲视频等，将所学理论概念与现实事件联系起来，能够促进学生的情感和认知联系[34]。例如通过对学术不端负面新闻事件剖析，使学生充分认识到无论是对于科研人员自身、还是整个科研团队、甚至于整个学术界，数据可重复都具有无可比拟的重要作用。⑤设计开放式问题鼓励学生思辩。帮助学生从反思和讨论中总结经验，为深入的创造性探索提供动力。针对数据伦理与意识部分，可将科研争议话题、已有研究存在的缺陷、提高可重复性所应采取的科学运作方式、道德限制等作为辩论主题[35]，引导学生进行广泛的批判性思考，通过讨论获得最佳认知。不仅能有效培养学生的批判性思维，还极大增强学习动力和参与度。

3.5 Re-DLE 评估评价

医学馆应积极创建可重复性数据素养能力的量化评估体系，用以精准测量用户的能力状况和追踪变化情况，为不断调整、改善教育目标和内容提供依据。例如华盛顿大学图书馆数据馆员参与开发的实证可重复性评估框架（RepeAT）[36]，加州大学旧金山分校图书馆建立的可重复性行为量化清单[37]，都能更好地了解目前生物医学研究人员在提高研究透明度、可及性、可重复方面的做法，为后续 Re-DLE 研讨会建立创造条件。同时，建立贯穿课前课后的教学效果评价机制，课前课后均通过问卷调查、半结构化访谈等方式获取用户能力、认知态度、行为方式等数据，并做详细分析对比，来评估教育对学生的积极作用、存在问题、学习阻力、改进方向，以便进一步指导和改善课堂教学。

4 Re-DLE 运行保障因素

4.1 加强队伍建设

医学馆必须加强具备教学胜任力的专业馆员队伍建设，形成 Re-DLE 运行的人力保障。定期开展针对性的教学能力培训，使馆员具备 Re-DLE 所需的知识理论储备和操作技能，以便能在教学中向学生提供更为专业的指导和反馈。以哈佛医学院为首的 6 所大学医学馆共同创建的科研数据管理馆员学院（Research Data Management Librarian Academy，RDMLA）提供免费课程，旨在帮助图书馆从业人员通过学习数据科学技能，制定可重复性数据标准、并学会如何与外界合作开发课程资源，将透明度和可重复性原则纳入定量研究培训和支持数据管理活动，来满足并支持科研用户日益增长的可重复研究实践需求[38]。此外，提升教学意愿是教师教学能力建设的源泉。应建立相关教

学激励机制，鼓励馆员开展教学实践。耶鲁大学等三校共建的可重复性数据策划（Data CuRe）培训计划项目通过提供津贴奖励和专业培训机会[39]，激发馆员对Re-DLE教育的兴趣和投入力度，提升教学能力。

4.2 夯实资源基础

教育资源是开展教育的基础与依托，也是学生开展培训与练习的必要条件。医学馆首先应多渠道整合各类的优质教材资源，开发在线教育资源库，为师生构筑完备的教育资源储备。美国大多数医学馆在开展Re-DLE过程中，仅有少数能够自主开发教材，大部分都借鉴了来自于美国国立卫生研究院（National Institutes of Health，简称NIH）的严谨性和可重复性培训模块资源、国立普通医学科学研究所（National Institute of General Medical Sciences，NIGMS）促进数据可重复性培训交流中心以及各级生物医学学会等已有教学培训资源，有效确保了资源权威性和规范性。其次，根据用户需求量身定制资源，通过预先调查掌握学生的水平现状，制定按专题领域分类的教学内容，结合所需的培训主题确定资源范围，根据每个主题的可用资源量和预计时间合并或拆分这些材料。最后，应加大资金投入力度，购入相关科研数据管理、开源软件平台、版本控制、自动分析、文学编程等软件工具。允许学生在稳定的平台上，实现发布、存储、共享科研过程中的软件、数据集、实验方案、工作流程和注释等，提高自身数据成果的可重复性；并支持学生利用云计算研究平台Code Ocean、应用容器引擎Docker等开展数据引用与复制练习。

4.3 推动教育交流

教育生态环境是对教育的产生、存在和发展起制约和调控作用的多元环境体系，创建一个知识高度交互共享的教育环境，有助于形成积极健康的教育生态。医学馆一方面要推动教育主体之间的交流，如组织学术论坛、成立教育联盟，促使教学者、管理者、协同者之间展开对话，凝聚共识，推动可重复与开放科学教育理念的快速传播，就提高教师育人能力、师资建设、资源开发、完善教学体系等话题展开更积极地讨论合作，为持续推动创新Re-DLE体系而共同努力。如美国佛罗里达大学图书馆、荷兰马斯特里赫特大学图书馆均组织开展了规模较大的专题教学研讨会，网罗志同道合的教育工作者分享包含Re元素的数据素养教学经验，共同探索教育创新途径。另一方面，推动教育受众群体之间的学习交流。欧洲众多医学馆近几年积极致力于强化开放研究社区的建设，如英国纽卡斯尔大学图书馆主导建立的开放科学期刊俱乐部（Reproducibilitea Journal Club）[40]，拥有固定交流场地和详细的课程安排，为本校科研群体提供了定期开展可重复和开放科学的主题交流渠道。而美国多数医学馆也建立了数据社区（Data Communities）、数据俱乐部（Data Club）等的跨学科用户交流平台，定期举办研讨会和社交活动，推动科研人员之间共同学习和处理数据，在项目进程中使自身的数据科学技能和开放与可重复原则得到有效融合。

4.4 完善制度建设

医学馆必须通过完善制度建设，为教育保驾护航。①对内建立专职馆员制度。专岗专职，全面负责Re-DLE教学组织与管理、教学资源筹备、用户调研等工作，随时响应教育的实际发展需要，使工作具有效率和针对性。例如佛罗里达大学图书馆设立的可重复馆员（Reproducibility Librarian）专职岗位，密歇根大学图书馆数据服务部与本校数据科学研究所共同设立的“数据管理和研究可重复性专家（Data Curation and Research Reproducibility Specialist）”的职位[41]，立足于满足本校教研群体的科研实践需求，负责就如何促进可重复研究提供资源导航、专业咨询和培训指导。②对外建立协同合作机制。生物医学科研的专业性、复杂性，以及Re-DLE教育内容内涵与外延的丰富程度，决定了医学馆仅凭一己之力难以承担，必须与基金资助机构、学术出版机构、社会组织、企业和学校等多方科研利益主体加强对话沟通以建立长期有效的合作关系，主导构建Re-DLE教育共同体，明确职责分工，发挥协作优势。例如针对不同教育主题模块，

DOI：10.13998/j.cnki.issn1002-1248.23-0158

可与具备对口专业优势的机构或人才合作来提升内容丰富程度。国外医学馆开展数据技能教育都争取到了来自软件工坊（Software Carpentry）、数据工坊（Data Carpentry）、图书馆工坊（Library Carpentry）等专业学习组织的教师支持，统计素养教育则邀请医学统计学、生物统计学、医学信息学的专业教学人才，数据伦理教育则会同本校的学术诚信委员会、科研管理部门展开工作。埃默里大学图书馆在筹备系列研讨会时，组建了馆员、校内教师、校外专家跨专业联合教学团队，确保了高质量的教学活动。

参考文献：

[1] GOODMAN S N, FANELLI D, IOANNIDIS J P A. What does research reproducibility mean?[J]. Science translational medicine, 2016, 8(341): eaaf5027.

[2] IOANNIDIS J P A. Why most published research findings are false[J]. PLoS medicine, 2005, 2(8): e124.

[3] MULLANE K, WILLIAMS M. Unknown unknowns in biomedical research: Does an inability to deal with ambiguity contribute to issues of irreproducibility?[J]. Biochemical pharmacology, 2015, 97 (2): 133–136.

[4] BEGLEY C G, IOANNIDIS J P A. Reproducibility in science[J]. Circulation research, 2015, 116(1): 116–126.

[5] PAPAMOKOS G V. The nature of the biological material and the irreproducibility problem in biomedical research[J]. The EMBO journal, 2019, 38(4): e101011.

[6] VOELKL B, WüRBEL H. A reaction norm perspective on reproducibility[J]. Theory in biosciences, 2021, 140(2): 169–176.

[7] SAMSA G, SAMSA L. A guide to reproducibility in preclinical research[J]. Academic medicine, 2019, 94(1): 47–52.

[8] DAKIN R, BINNING S A. Slow improvement to the archiving quality of open datasets shared by researchers in ecology and evolution[J]. Proceedings of the royal society B: Biological sciences, 2022, 289 (1975): 1–8.

[9] SAMUEL S, KÖNIG –RIES B. Understanding experiments and research practices for reproducibility: An exploratory study[J]. PeerJ, 2021, 9: e11140.

[10] YU B, HU X A. Toward training and assessing reproducible data analysis in data science education[J]. Data intelligence, 2019, 1(4): 381–392.

[11] KARATHANASIS N, HWANG D, HENG V, et al. Reproducibility efforts as a teaching tool: A pilot study[J]. PLoS computational biology, 2022, 18(11): e1010615.

[12] HUPPENKOTHEN D, ARENDT A, HOGG D W, et al. Hack weeks as a model for data science education and collaboration[J]. Proceedings of the national academy of sciences of the United States of America, 2018, 115(36): 8872–8877.

[13] 叶新友,张路路,孔成果,等.国内研究生科学数据素养能力评价及高校图书馆培养体系构建研究[J].农业图书情报学报,2021,33 (11):63–73.
YE X Y, ZHANG L L, KONG C G, et al. Evaluation of scientific data literacy competency for postgraduates in China and construction of data literacy education system[J]. Journal of library and information science in agriculture, 2021, 33(11): 63–73.

[14] ALVES V M, AUERBACH S S, KLEINSTREUER N, et al. Curated data in – Trustworthy In silico models out: The impact of data quality on the reliability of artificial intelligence models as alternatives to animal testing[J]. Alternatives to laboratory animals, 2021, 49(3): 73–82.

[15] LEIPZIG J, NÜST D, HOYT C T, et al. The role of metadata in reproducible computational research[J]. Patterns, 2021, 2(9): 100322.

[16] DIRNAGL U. Rethinking research reproducibility[J]. The EMBO journal, 2019, 38(2): e101117.

[17] ELLIS R J. Questionable research practices, low statistical power, and other obstacles to replicability: Why preclinical neuroscience research would benefit from registered reports[J]. eneuro, 2022, 9 (4): ENEURO.0017–22.2022.

[18] National Academies of Sciences E, Engineering, Medicine, et al. Reproducibility and replicability in science[M]. Washington, DC: National Academies Press, 2019.

[19] STECKLER T. Editorial: Preclinical data reproducibility for R&D – The challenge for neuroscience[J]. SpringerPlus, 2015, 4(1): 1–4.

[20] FRANÇA T F, MONSERRAT J M. Reproducibility crisis in science or unrealistic expectations?[J]. EMBO reports, 2018, 19(6): e46008.

[21] GRÜNING B, CHILTON J, KOSTER J, et al. Practical computational reproducibility in the life sciences[J]. Cell systems, 2018, 6(6): 631–635.

[22] ALTER G, GONZALEZ R. Responsible practices for data sharing[J]. American psychologist, 2018, 73(2): 146–156.

[23] BEZUIDENHOUT L, QUICK R, SHANAHAN H. "Ethics when you least expect it": A modular approach to short course data ethics instruction[J]. Science and engineering ethics, 2020, 26(4): 2189–2213.

[24] BRITO J J, LI J, MOORE J H, et al. Recommendations to enhance rigor and reproducibility in biomedical research[J]. GigaScience, 2020, 9(6): giaa056.

[25] COLLINS F S, TABAK L A. Policy: NIH plans to enhance reproducibility[J]. Nature, 2014, 505(7485): 612–613.

[26] Clearinghouse for training modules to enhance data reproducibility[EB/OL]. [2023 –04 –15].https://www.nigms.nih.gov/training/pages/clearinghouse–for–training–modules–to–enhance–data–reproducibility.aspx.

[27] Training and Other Resources | Grants.nih.gov[EB/OL]. [2023–04–15]. https://grants.nih.gov/policy/reproducibility/training.htm.

[28] BOSCH G, CASADEVALL A. Graduate biomedical science education needs a new philosophy[J]. mBio, 2017, 8(6): e01539–17.

[29] Preparing FAIR data for reuse and reproducibility | Research Data Management Service Group[EB/OL]. [2023–04–15]. https://data.research.cornell.edu/content/preparing–fair–data–reuse–and–reproducibility.

[30] 孔祥辉, 陈卓. 美国医学院校图书馆开展可重复性研究教育的调查与分析[J]. 数字图书馆论坛, 2023, 19(1): 26–33.
KONG X H, CHEN Z. Investigation and analysis of reproducibility research education in U.S. medical libraries[J]. Digital library forum, 2023, 19(1): 26–33.

[31] 古婷骅, 王腾利, 方舟. 大数据背景下高职院校数据素养教育研究——基于广东省科技干部学院商科学生的抽样调查[J]. 农业图书情报学报, 2021, 33(1): 80–91.
GU T H, WANG TL, FANG Z. Research on data literacy education in vocational colleges and universities: A survey on students majoring in business in Guangdong polytechnic of science and technology[J]. Journal of library and information science in agriculture, 2021, 33(1): 80–91.

[32] KARATHANASIS N, HWANG D, HENG V, et al. Reproducibility efforts as a teaching tool: A pilot study[J]. PLoS computational biology, 2022, 18(11): e1010615.

[33] ReproHack – March 2021[EB/OL]. [2023–03–09]. https://www.imperial.ac.uk/computational–methods/rse/events/reprohack–mar21/.

[34] Teaching reproducible research for medical students and postgraduate pharmaceutical scientists | BMC Research Notes[EB/OL]. [2023–03–09].https://bmcresnotes.biomedcentral.com/articles/10.1186/s13104–021–05862–8.

[35] Reproducible Analysis | Colorado State University Libraries[EB/OL]. [2023 –03 –09]. https://lib.colostate.edu/services/data –management/reproducible–analysis/.

[36] MCINTOSH L D, JUEHNE A, VITALE C R H, et al. Repeat: A framework to assess empirical reproducibility in biomedical research[J]. BMC medical research methodology, 2017, 17(1): 1–9.

[37] DEARDORFF A. Assessing the impact of introductory programming workshops on the computational reproducibility of biomedical workflows[J]. PLoS One, 2020, 15(7): e0230697.

[38] RDMLA: Free training course for librarians now offers the opportunity to earn CE credits [EB/OL]. [2023–04–17]. https://www.elsevier.com/connect/library–connect/rdmla–free–training–course–for–librarians–now–offers–the–opportunity–to–earn–ce–credits.

[39] CURE Training | Curating for Reproducibility [EB/OL]. [2023–04–13]. https://curating4reproducibility.org/training/.

[40] ReproducibiliTea | University Library | Newcastle University[EB/OL]. [2023 –05 –07]. https://www.ncl.ac.uk/library/academics –and –researchers/research/open–research/reproducibilitea/.

[41] Data Curation and Research Reproducibility Specialist [EB/OL]. [2023 –04 –14].https://careers.umich.edu/job_detail/227704/data –curation–and–research–reproducibility–specialist.

研究论文
DOI：10.13998/j.cnki.issn1002-1248.23-0158

Medical Data Literacy Education System in Reproducibility Crisis

KONG Xianghui, SUN Pu*
(Jinzhou Medical Universit, Jinzhou 121017)

Abstract: [Purpose/Significance] The biomedical research field is suffering from reproducibility crisis, which has become one of important issues under the background of the rise of the data-intensive research paradigm. As one of the most important attributes of scientific research by empirical data study,reproducibility needs to be improved by good data practices of researchers. How to effectively improve the data literacy of researchers has become the key point to solve the crisis. However, the relevant research is basically in the blank condition. The paper aims to establish a new data literacy education system for reproducibility crisis, in order to fill the current research gap and provide reference for implementing the relevant education in our country. [Method/Process] Firstly, the paper clarifies the relationship between reproducibility crisis and data literacy by using content analysis: the inappropriate data behavior of researchers may bring serious problems in many respects, such as research data, methods, process, environments and results, which could eventually lead to the irreproducible research. Then, we redefine the concept of data literacy education. Secondly, based on the summarization of the existing foreign research results and practice, the paper builds the Reproducibility Data Literacy Education (Re-DLE) system from the perspective of educational goals and content, subjects and objects, teaching methods, implementation strategies, and evaluation. At last, it proposes the necessary guarantee factors for the operation of the system. [Results/Conclusions] The ultimate goal of Re-DLE is to improve research reproducibility, bulid the educational content framework on the theory of data life cycle, and divide the main content into three dimensions: re-data awareness, re-data skills, and re-data ethics, each of which includes some clear educational objectives, subject modules and detailed instructions Medical libraries have a wealth of teaching experience and should become the educational main body for the broader biomedical research community. the establishment of diversified training methods, diversified teaching strategies and evaluation methods, in other words, we need to strengthen the team building of teaching librarians, consolidate the educational resources foundation, promote educational exchanges, and improve the internal and external cooperative system, so as to push forward the building of the Re-DLE system. The research results of this paper not only can be seen as a theoretical breakthrough, but also provide the theory basis for the development and implementation of education. In addition, due to the limitation of methods, the paper can be used as a qualitative research, which still has some problems to be solved. In the future work, we need to build more scientific and effective Re-DLE system by using empirical research methods.

Keywords: data literacy education; data management; reproducibility; medical informatics; medical library