

基于系统性文献综述的医学科学数据质量评价指标框架构建

冉尤冰 胡志民

(中国医学科学院/北京协和医学院卫生健康管理政策学院 北京 100730)

〔摘要〕 **目的/意义** 构建医学科学数据质量评价指标框架, 为医学科学数据高质量建设和医学标准规范体系完善提供参考。**方法/过程** 通过系统性文献综述, 总结科学数据质量评价维度和通用指标, 结合医学科学数据特征构建评价指标框架。**结果/结论** 该框架包含 5 个维度和 13 项指标, 引入科学数据生命周期理论划分维度, 探索性纳入伦理审查和匿名化指标, 更契合医学科学数据高质量发展实际需求。

〔关键词〕 医学科学数据; 质量评价; 框架构建; 数据生命周期

〔中图分类号〕 R-058 **〔文献标识码〕** A **〔DOI〕** 10.3969/j.issn.1673-6036.2025.07.002

Construction of Quality Evaluation Indicator Framework for Medical Scientific Data Based on Systematic Literature Review

RAN Youbing, HU Zhimin

School of Health Policy and Management, Chinese Academy of Medical Science & Peking Union Medical College, Beijing 100730, China

〔Abstract〕 **Purpose/Significance** To construct a quality evaluation indicator framework for medical scientific data, providing a reference for enhancing high-quality medical scientific data construction and refining medical standardization systems. **Method/Process** A systematic literature review is conducted to synthesize quality evaluation dimensions and generic indicators for scientific data. The framework is then constructed by integrating characteristics specific to medical scientific data. **Result/Conclusion** This framework consists of 5 dimensions and 13 indicators, introduces the theory of scientific data lifecycle as a criterion for dividing dimensions, and explores the inclusion of ethical review and anonymization indicators, which is more in line with the practical needs of medical scientific data high-quality construction.

〔Keywords〕 medical scientific data; quality evaluation; framework construction; data life cycle

〔修回日期〕 2025-07-06

〔作者简介〕 冉尤冰, 硕士研究生; 通信作者: 胡志民, 博士, 研究员, 硕士生导师。

〔基金项目〕 国家科技重大专项项目 (项目编号: 2023ZD0509701); 中国医学科学院医学与健康科技创新工程项目 (项目编号: 2021-I2M-1-057)。

1 引言

作为国家科技创新体系的基础性战略资源, 科学数据驱动学科发展和社会经济转型。针对其利用率不足及跨境流失问题, 我国于 2008 年启动数据汇交制度建设, 2018 年通过《科学数据管理办法》, 将其升级为财政资助项目结项的强制性前置条件。在数字经济时代, 高质量科学数据已成为驱动科技

创新的核心生产要素。《信息技术 数据质量评价指标》(GB/T 36344—2018)指出,数据质量是“在指定条件下使用时,数据特性满足明确和隐含要求的程度”^[1]。相似地,本研究将科学数据质量定义为“科学数据特性满足用户所从事的研究活动要求的程度或满足研究期望的程度”。

医学科学数据具有较高研究价值、社会价值和经济价值^[2],高质量数据集建设已成为促进我国医疗行业数字化转型的关键举措^[3]。然而,目前我国尚未建立医学科学数据质量评价标准规范,各医疗机构数据存储结构不一致,异质性强,阻碍数据开放共享,亟待建立严格质量控制体系。

数据质量评价指标框架是医学科学数据质量控制的机制,也是医学科学数据标准体系的重要组成部分。但当前国内研究主要聚焦平台建设^[4]和共享机制^[5],针对评价指标框架的研究较少。鉴于此,本研究采用“从一般到特殊”的思路,通过系统性文献综述总结通用科学数据质量评价指标框架,再针对医学科学数据特征调整完善。

2 研究方法

系统性文献综述是利用预先设计的明确过程对文献进行全面评述的方法^[6],可增强研究严谨性、可重复性并降低主观偏倚。研究过程遵循《Cochrane 干预措施系统评价手册》中的推荐步骤,并参照《系统综述和元分析优先报告的条目》(preferred reporting items for systematic reviews and meta-analyses, PRISMA)报告研究结果。

2.1 文献纳排标准

纳入标准:研究对象为“科学数据 (scientific data/research data)”；研究主题为科学数据质量评价指标；提出的质量评价指标成体系或框架。排除标准:非原始研究,如综述、声明；专利或数据集出版物；未报告质量评价指标或仅应用现有指标；质量评价对象非科学数据；与质量评价无关的研究。

2.2 文献检索及筛选

以“科学数据质量评价”为主题,系统检索中国知网、SinoMed、Web of Science 和 Science Direct 数据库自建库至 2024 年 12 月 31 日的出版物。在文献回顾和科学数据领域专家指导下,以“科学数据”“研究数据”“数据质量”“数据绩效”“评估”“评价”为中文检索词,以“scientific data”“research data”“data quality”“quality evaluation”“quality assessment”为英文检索词,结合布尔逻辑运算符构建检索式。

两名作者根据检索策略同步检索,共检索出 517 篇文献。首先排除重复文献,通过机器和人工去除 33 篇；其次阅读文献标题和摘要并进行初筛,根据纳入标准排除 439 篇；再次阅读全文复筛,根据排除标准排除 27 篇；最后检索复筛中排除的 27 篇文章的参考文献和引证文献,补充纳入两篇符合纳排标准的文献,构成最终文献集。为避免主观选择偏差,上述步骤由两名作者独立完成。如果发生意见不一致,则通过协商达成共识。最终文献集共纳入 20 篇研究科学数据质量评价指标的文献,其中 9 篇聚焦通用领域^[7-15],6 篇聚焦医学领域^[16-21],5 篇聚焦其他特定领域^[22-26],见图 1。

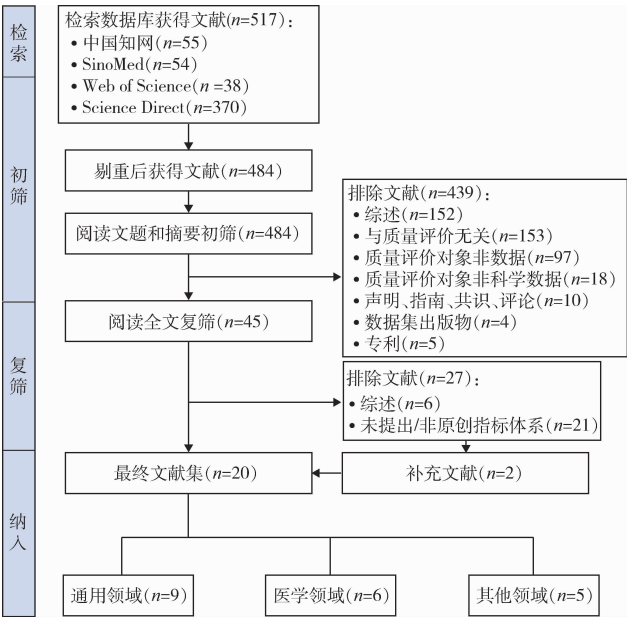


图 1 文献筛选流程

2.3 数据编码

针对纳入文献中以图表展示的评价体系，采用 3 级编码结合频次统计进行数据提取与分析。首先提取质量评价维度和指标，比较分析原始文本，提炼关键特征并统一命名。如“真实性”“可信性”“客观性”强调数据来源真实，“可靠性”“可溯源性”“可重复性”要求数值稳定，二者高度关联，因此统一编码为“真实可靠性”，此类编码视作一级编码。其次基于一级编码结果进行频次统计，保留频次大于等于 2 的一级编码，通过聚类与关联分析归纳核心类别。最后排除频次低于 10 的指标，得到通用指标。编码由两名作者同步实施，分歧通过协商解决。

3 通用科学数据质量评价指标

3.1 纳入文献质量

本研究纳入期刊文献 20 篇，来源于 18 种期刊（中文 11 种，英文 7 种）。13 篇中文文献中，3 篇刊载于中文社会科学引文索引（Chinese social sciences citation index, CSSCI）与《中文核心期刊要目总览》（又称北大核心期刊）双收录期刊，3 篇

刊载于北大核心期刊，1 篇刊载于 CSSCI 扩展版期刊；被引频次方面，3 篇高于 20 次，4 篇 10 ~ 20 次，6 篇低于 10 次。7 篇英文文献中，5 篇发表于期刊引证报告（journal citation reports, JCR）Q2 期刊；被引频次方面，2 篇高于 50 次，2 篇 10 ~ 50 次，3 篇低于 10 次。

3.2 科学数据质量评价维度

“数据质量维度”指一组数据质量属性的集合，这些属性共同表征数据质量的某个特定方面或概念^[27]。信息提取显示，15 篇明确质量评价维度的文献呈现 4 类划分逻辑：一是基于数据生命周期阶段性特征^[7-10]；二是依据项目实施流程^[16-17,19]；三是聚焦关键质量指标^[18,20-22]；四是解构数据内在结构^[11-13,15]。针对前两类，本研究基于科学数据生命周期管理模型^[28]对原始文本进行一级编码；针对后两类，综合相关文献提出的质量评价维度，将原始文本编码为数据内容、形式、效用、访问权限等核心类别。

编码结果显示，科学数据质量评价维度可归纳为动态与静态两种分析视角，见表 1。15 篇文献中，7 篇（46.7%）采用动态视角，8 篇（53.3%）采用静态视角，分布较为均衡。

表 1 科学数据质量评价维度

视角（二级编码）	质量评价维度（一级编码）	原始文本举例	文献频次
动态视角下的质量评价维度	数据收集阶段	“数据采集”“data construction”	6
	数据处理阶段	“数据分析与加工管理”“数据处理质量、数据分析质量”	6
	数据存储阶段	“数据存储和管理”“data operation; during regular updates and maintenance, the consistency, accuracy, and timeliness of data must be maintained”	5
	数据利用阶段	“数据服务”“data utilization”	5
	计划制定阶段	“制定数据管理计划”“临床试验启动阶段”	4
静态视角下的质量评价维度	数据内容质量	“研究数据固有的内在质量”“completeness”	8
	数据形式质量	“研究数据表达形式的质量”“consistency”	7
	数据效用质量	“数据利用质量”“quality indicator of platform construction is consisted of two parts: one is web site content and another is operation effect”	5
	数据访问权限	“数据访问质量”“quality indicators on the resource development”	3

动态视角下，科学数据质量评价维度强调全生命周期质量控制，涵盖数据计划制定、收集、处

理、存储以及利用 5 大阶段。7 篇动态视角文献中，数据收集和处理维度的关注度最高，数据存储和利用维度次之，计划制定维度研究相对薄弱。静态视角采用截面分析，假设数据资源及其环境稳定，主要评估数据内容、形式、效用和访问权限 4 个维度。其中，数据内容与形式为核心维度，数据效用和访问权限维度关注度稍低。这表明静态视角更侧重数据本身属性和显性特征，而数据效用与访问权限作为共享效果的重要指标，研究深度有待提升。综上，研究成果揭示了不同视角下科学数据质量评价维度的差异。从应用场景看，动态视角适用于科

研项目管理和数据共享平台运营，通过全流程控制保障数据质量；静态视角则在数据汇交和交易场景中优势显著。

3.3 科学数据质量评价通用指标

编码分析共得到 16 项质量评价指标（文献频次大于等于 2）。为明确指标内容与重点，采用静态视角进行二级编码，见表 2。排除文献频次低于 10 的指标后，筛选出完整性、准确性、时效性等 9 项通用指标，见表 3。

表 2 科学数据质量评价指标

类别（二级编码）	质量评价指标（一级编码）	原始文本举例	文献频次
数据内容质量	完整性	“全面性” “missing values”	17
	准确性	“准确性” “accuracy”	15
	时效性	“更新及时性” “有效性” “timeliness”	11
	真实可靠性	“所产出信息的客观性” “reliability”	11
	易理解性	“可理解性” “易理解性”	6
	相关性	“内容关联性” “relational conformance”	5
	可信性	“所产出信息的可信性” “可信度”	2
数据形式质量	一致性	“一致性” “consistency”	11
	规范性	“标准化” “data record mismatch”	11
	机器可读性	“机器可读性” “易读取性”	4
数据访问权限	可访问性	“公开性” “web site content”	10
	可获取性	“可获得性” “dataset service development indicator”	10
	可存储性	“安全性” “数据备份”	7
数据效用质量	可重用性	“可重用性” “增值性”	10
	数据利用率	“利用率” “utilization of GPS observations”	4
	数据影响力	“数据影响力” “operation effect indicator”	3

表 3 科学数据质量评价通用指标

质量评价维度	质量评价指标	指标描述
数据内容质量	完整性	数据实体、元数据以及其他所需信息的完整程度
	准确性	数据内容反映实际观察情况的准确程度以及数值记录过程的精确程度
	时效性	数据资源在特定时期内反映的事物状态、趋势、特征与现实的相符程度
	真实可靠性	数据来源的真实性和数值的稳定性
数据形式质量	一致性	数据元素在语法结构和表达上的统一
	规范性	数据资源符合现行标准、行业规范、领域共识及平台规章的程度
数据访问权限	可访问性	访问数据资源的可行性和便利性
	可获取性	下载或在线分析数据资源的可行性与操作便捷性
数据效用质量	可重用性	数据资源在重用过程中创造新成果的潜力

3.3.1 数据内容质量 是对数据资源语义层面的多维评价。目前普遍采用完整性、准确性、时效性和真实可靠性进行评价。在数据收集阶段，完整性、准确性和真实可靠性易受人为输入误差、数据缺失及冗余记录影响。完整性可通过设定期望数据缺失比例、元数据标准及规范来评估；准确性可结合异常值、类内差异与理论预测值偏差等指标来评价；真实可靠性要求研究具有可重复性，以证明数据真实存在、数值稳定可信，可通过核查原始数据日志、伦理审查文件及研究者认证签名等记录进行初步评估；时效性强调数据资源与时间的关系，随时间推移逐渐降低，过时数据可能无法反映现实，须通过周期性更新维持。

3.3.2 数据形式质量 是对数据资源语法层面的全面评价，关注数据描述的规范与统一。一致性和规范性是常用评价指标。一致性体现在 3 个方面：格式一致，包括字段类型、编码方式及存储结构标准化；逻辑一致，数据内在关联符合预设计算逻辑；更新一致，保障数据在同一时空的版本和状态一致。规范性涉及数据格式、命名规则及元数据管理规范的遵循情况。为保证数据一致性和规范性，数据收集前应制定元数据标准及数据字典。

3.3.3 数据访问权限 衡量用户访问和获取平台存储数据的权限，反映科学数据开放共享的水平，影响其长期成效。常用评价指标为可访问性和可获取性。可访问性关注访问权限与渠道，受数据分类分级标准与平台技术水平影响，表现为：在保障数据安全和隐私前提下，通过身份认证、权限分级等技术实现差异化访问控制。可获取性衡量开放共享实践效果，须综合考量接口开放水平、在线计算资源的可及性等技术参数，以及知识产权协议与数据使用条款等制度约束。对于“可用不可见”的特殊数据资源，评价应着重考察元数据的可访问性及在线分析工具的易用性。

3.3.4 数据效用质量 衡量数据资源使用价值水平，包括学术研究价值与应用转化价值，核心评价指标为可重用性。可重用性反映数据资源的再利用价值，随着数据利用程度加深而降低。可重用性受数据资源的独特性和数据内容的冗余性影响，独特

性决定再利用价值的上限，冗余性则影响再利用阶段投入的数据处理成本。

4 医学科学数据质量评价指标框架

4.1 医学科学数据特征

医学科学数据具有区别于其他学科数据的显著特征。一是隐私保护要求高。医学研究涉及个人健康信息，共享时要严格去标识化处理，通过技术手段确保信息不可溯源，最小化隐私泄漏风险。二是伦理原则全程约束。伦理审批是医学数据生产的特殊环节，是保障患者权益和规范研究行为的必然要求。医学科学数据再利用应符合伦理原则。三是数据多源异构。医学科学数据来源包括临床试验、诊疗、监测及社会统计；形态涵盖影像与非影像。此特性导致数据整合复杂，标准化处理是其高质量发展的关键。

4.2 医学科学数据质量评价指标框架

本研究在通用指标基础上，结合医学科学数据特征，构建包含 13 项指标的医学科学数据质量评价指标框架，见表 4。该框架采用动态视角划分评价维度，依据数据全生命周期各阶段质量控制要点，分阶段设置评价指标。其中，科学数据管理计划（data management plan, DMP）是描述项目全生命周期数据管理、处理与共享的规划文件^[29]，为后续数据质量控制提供依据。DMP 须明确元数据标准、数据共享权限、管理责任主体等核心要素，制定时应重点评估细粒度、规范性与价值性^[8,10]。伦理审查是指涉及科技伦理敏感领域的研究须按《科技伦理审查办法（试行）》申请审查。项目单位科技伦理委员会出具的“同意”意见文件，通常视为数据符合伦理规范。评估待核查文件真实性及项目活动是否超出批准范围；存在虚假或超范围情况的，相关数据禁止流通利用。机器可读性要求数据处理阶段须实现数据可互操作。可互操作性强调跨系统兼容性，涉及数据表达规范与机器可读格式两个维度。采用形式化知识表示语言可消除语义歧义，提升机器可读性。数据加工应采用易于机器解

析的格式，如 JSON、RDF、XML，以支持跨平台共享。匿名化要求个人身份信息不可识别，其原理是通过移除或混淆敏感属性，降低重识别风险并尽量保留数据价值^[30]。常见技术包括 k-匿名、差分隐私、同态加密等，其中同态加密常用于“可用不可见”数据共享。需注意，现有技术无法实现完全匿名化，隐私泄漏风险依然存在。

表 4 医学科学数据质量评价指标框架

质量评价维度	质量评价指标
计划制定阶段	DMP 制定 ^[8,10] 、伦理审查
数据收集阶段	完整性 ^[12,15] 、准确性 ^[13,16] 、真实可靠性 ^[10,12] 、一致性 ^[21] 、规范性 ^[9,16]
数据处理阶段	机器可读性 ^[8,11] 、匿名化
数据存储阶段	时效性 ^[12] 、可访问性 ^[15]
数据利用阶段	可获取性 ^[12,18] 、可重用性 ^[11]

5 讨论

5.1 评价指标框架的内容分析

现有医学科学数据质量评价研究多聚焦特定数据类型指标框架构建，如电子病历数据^[19]、临床试验数据^[16]、真实世界研究数据^[17]。针对单一数据类型框架应用受限、推广成本高的问题，本研究在 9 项通用指标基础上，结合医学数据特征构建评价指标框架，扩展了其适用范围。与现有框架^[16-21]相比，本框架在评价维度方面，引入科学数据生命周期理论，系统性识别各阶段质量控制关键节点，支持质量问题精准追溯与及时解决，有利于建立“评价-反馈-优化”闭环机制，减少后期质量问题导致的返工和重复核验；评价指标方面，重视开放共享中的伦理与隐私问题，探索性纳入伦理审查与匿名化指标，强调“合法合规优先”原则，为共享环境下的数据安全与伦理风险防控提供保障。

5.2 评价指标框架的应用场景分析

5.2.1 医学科研团队 医学科研团队可通过 DMP 制定和伦理审查指标评价优化课题计划。一方面，涉及患者身份和健康信息的数据收集必须通过伦理

审查；另一方面，完善的 DMP 既能保障科研全流程数据质量，又能向资助方展示可行方案。在数据收集阶段，须通过 DMP 评价数据的完整性、准确性、真实可靠性、一致性、规范性。以虚拟高血压数据集为例：李四的收缩压和舒张压缺失（完整性）；王五的舒张压异常（准确性）；可溯源的患者信息保障真实可靠性；数据格式统一体现一致性；纳入高血压关键指标（收缩压/舒张压）符合现行标准（规范性），见表 5。团队应根据评价结果及时采取措施，如重新收集数据。

表 5 高血压数据（虚拟）

患者姓名	年龄 (岁)	性别	婚姻状况	收缩压 (mmHg)	舒张压 (mmHg)
张三	58	男	未婚	157	79
李四	67	女	丧偶	—	—
王五	62	男	已婚	180	10
.....					

在数据处理阶段，须将数据转换为易于机器解析的格式，方便后续数据分析，如将文本信息编码为二分类变量；匿名化处理患者隐私信息，如以编号代称患者姓名。在数据存储阶段，评价时效性可激励科研团队及时分析数据，高时效性数据可能加速成果产出；评价可访问性要求团队恰当设置访问权限，防止数据被篡改。在数据利用阶段，可获取性要求设置数据下载权限，防止核心数据泄漏；可重用性评估数据再利用价值，是数据共享或交易的关键考量因素。

5.2.2 医学科学数据中心 在计划制定阶段，科学数据中心通过前置审查科研团队的 DMP，辅助科研团队完善 DMP，以提高汇交数据的质量；通过审查伦理文件，进行伦理风险防控，以确保汇交数据的合规性。在数据汇交时，以完整性、准确性等指标作为项目结项标准，筛选具有开放共享价值的数据。对于拟开放共享的数据，评价其机器可读性和匿名化程度，作为数据发布的关键标准。前者影响用户数据分析的可操作性，后者满足《个人信息保护法》与《数据安全法》要求。在数据存储阶段，实施基于敏感度的分级访问制度保障核心数据安

全, 提供数据字典等辅助工具, 对低时效性数据更新或归档。在数据利用阶段, 确保发布数据的可获取性, 支持下载或在线分析, 重点推荐高可重用性数据, 如长期重症监护室数据、健康统计数据等。

当前医学科学数据质量评价缺乏统一标准, 各科研项目和科学数据中心各自探索质量评价指标, 难以广泛适用于其他医学科学数据。本框架可在医学领域广泛应用, 为制定行业或国家标准规范提供重要参考。建议基于科学数据生命周期理论构建标准规范, 将“合法合规性”作为基本要求和重要内涵纳入评价体系。

5.3 评价指标框架的局限性分析

本框架虽在医学领域有广泛适用性, 但仍存在实践局限。一方面, 医学科研团队与科学数据中心需求差异导致评价指标框架的权重配置侧重不同, 须紧密结合具体场景和目的, 采用层次分析法、熵权法等细化评价标准。因此本框架暂未配置各指标权重, 难以直接投入实践应用。另一方面, 本研究侧重理论构建, 缺乏实证检验, 有待进一步验证实际效用。

6 结语

本研究通过系统文献综述构建覆盖数据生命周期的医学科学数据质量评价指标框架, 为质量评价实践提供理论支撑, 助力医学科学数据高质量发展。作为阶段性成果, 后续研究将聚焦于指标权重配置及应用检验。未来研究在应用此框架时, 须采用德尔菲法完善评价标准细则, 科学配置指标权重, 并开展实证研究, 以增强其实用性。

作者贡献: 冉尤冰负责文献筛选与编码、论文撰写; 胡志民负责研究设计、文献筛选与编码、论文审核与修订。

利益声明: 所有作者均声明不存在利益冲突。

参考文献

1 信息技术 数据质量评价指标: GB/T 36344—2018 [EB/OL]. [2025-03-15]. <https://openstd.samr.gov.cn/>

bzgk/gb/newGbInfo? hcno = D12140EDFD3967960F51BD1A05645FE7.

- 2 万佳林, 贾晓峰, 胡志民. 基于多案例研究的生物医学科学数据开放共享策略分析 [J]. 医学信息学杂志, 2024, 45 (2): 20-25.
- 3 胡坚波. 加快建设高质量数据集 推动人工智能赋能行业发展 [EB/OL]. [2025-06-15]. https://www.nda.gov.cn/sjj/zwgk/zjjd/0303/20250303165949956429831_pc.html.
- 4 吴思竹, 王安然, 修晓蕾, 等. 欧美生物医学科学数据中心建设及启示 [J]. 数字图书馆论坛, 2022 (4): 2-10.
- 5 关键. 法律框架下构建医学科学数据共享机制和保障体系 [J]. 中国科技资源导刊, 2020, 52 (2): 62-68.
- 6 马鑫, 王芳. 元宇宙的概念、技术、应用与影响——一项系统性文献综述 [J]. 图书情报工作, 2023, 67 (18): 113-128.
- 7 胡良霖. 科学数据资源的质量控制和评估 [J]. 科研信息化技术与应用, 2009 (1): 50-55.
- 8 江洪, 王春晓. 基于科学数据生命周期管理阶段的科学数据质量评价体系构建研究 [J]. 图书情报工作, 2020, 64 (10): 19-27.
- 9 孔丽华, 习妍, 郎杨琴, 等. 数据期刊中科学数据的同行评议方法研究 [J]. 编辑学报, 2019, 31 (3): 262-266.
- 10 刘琼, 聂云贝, 刘桂锋. 科学数据质量评价指标体系构建及用户满意度研究 [J]. 图书情报导刊, 2022, 7 (12): 31-38.
- 11 孟银涛, 赵蕾霞, 于倩倩. 面向科学数据保存的科学数据评价指标体系研究 [J]. 农业图书情报学报, 2021, 33 (12): 48-59.
- 12 撒旭, 王健, 范智萱, 等. 数据期刊同行评议视角下科学数据质量评价指标识别 [J]. 图书情报工作, 2020, 64 (17): 123-130.
- 13 王志强, 杨青海. 科学数据质量及其标准化研究 [J]. 标准科学, 2019 (3): 25-30.
- 14 张文德, 陈汐, 包雪清, 等. 科研人员判断科学数据质量和权威性的行为模式研究 [J]. 数字图书馆论坛, 2023, 19 (9): 11-19.
- 15 周洁. 研究数据的质量评价指标体系研究 [J]. 图书情报导刊, 2019, 4 (12): 71-76.
- 16 陈朝华, 黄钦, 邓亚中, 等. 临床试验数据管理质量评价指标体系 [J]. 药学学报, 2015, 50 (11): 1374-1379.
- 17 尹文俊, 谢悦良, 黄玉洁, 等. 药品临床综合评价中真实世界数据质量控制的探讨和工作经验分享 [J]. 中国

- 药学杂志, 2024, 59 (4): 369–374.
- 18 HAIYAN L, MENG C. Quality assessment and indicator system frame of national medical scientific data sharing project [J]. World science and technology, 2009, 11 (4): 578–581.
- 19 LEE S, ROH G H, KIM J Y, et al. Effective data quality management for electronic medical record data using SMART DATA [J]. International journal of medical informatics, 2023, 180 (12): 105262.
- 20 KAHN M G, CALLAHAN T J, BARNARD J, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data [J]. eGEMs, 2016, 4 (1): 1244.
- 21 SCHMIDT C O, STRUCKMANN S, ENZENBACH C, et al. Facilitating harmonized data quality assessments. A data quality framework for observational health research data collections with software implementations in R [J]. BMC medical research methodology, 2021, 21 (1): 63.
- 22 王卷乐, 陈沈斌. 地学栅格格网数据质量评价指标与方法 [J]. 测绘科学, 2006 (5): 83–85, 82.
- 23 吴贤玲, 王秀珍, 黄敬峰. 面向水稻生长模拟的网络共享数据综合评价 [J]. 杭州师范大学学报 (自然科学版), 2012, 11 (5): 464–469.
- 24 FUCHS S, NORDEN B, NEUMANN F, et al. Quality – assurance of heat – flow data: the new structure and evaluation scheme of the IHFC global heat flow database [J]. Tectonophysics, 2023, 863 (9): 229976.
- 25 KONG Q, CHEN Y, FANG W, et al. Analysis of space – borne GPS data quality and evaluation of precise orbit determination for COSMIC – 2 mission based on reduced dynamic method [J]. Remote sensing, 2022, 14 (15): 3544.
- 26 OLSEN A R, VOLDNER E C, BIGELOW D S, et al. Unified wet deposition data summaries for North America: data summary procedures and results for 1980–1986 [J]. Atmospheric environment, 1990, 24 (3): 661–672.
- 27 WANG R Y, STRONG D M. Beyond accuracy: what data quality means to data consumers [J]. Journal of management information systems, 1996, 12 (4): 5–33.
- 28 张洋, 肖燕珠. 生命周期视角下《科学数据管理办法》解读及其启示 [J]. 图书馆学研究, 2019 (15): 37–43, 13.
- 29 殷沈琴, 徐萍. 国际数据管理计划工具比较研究 [J]. 图书馆杂志, 2024, 43 (4): 108–117, 142.
- 30 陈美, 梁乙凯. 开放政府数据隐私风险控制中个人数据匿名化研究 [J]. 图书馆学研究, 2021 (11): 66–71.

(上接第 6 页)

- 20 NIH Common Fund. Bridge2AI: propelling biomedical research with artificial intelligence [EB/OL]. [2024–12–27]. <https://bridge2ai.org/>.
- 21 NIH. The NIH’s artificial intelligence/machine learning consortium to advance health equity and researcher diversity program – data and research core [EB/OL]. [2024–12–27]. <https://www.aim-ahead.net/data-and-research-core/>.
- 22 NIH. The NIH’s artificial intelligence/machine learning consortium to advance health equity and researcher diversity program – aim [EB/OL]. [2024–12–27]. <https://www.aim-ahead.net/programs/aim-ahead-dicb-program/>.
- 23 NIH Office of Data Science Strategy. Advancing health research through ethical, multimodal AI [EB/OL]. [2025–01–02]. <https://datascience.nih.gov/sites/default/files/MAI-Solicitation-outline.pdf>.
- 24 CLARK T, CAUFIELD H, PARKER J A, et al. AI – readiness for biomedical data: Bridge2AI recommendations [EB/OL]. [2025–01–02]. <https://pubmed.ncbi.nlm.nih.gov/39484409/>.
- 25 赵国屏. 生物医学研究范式的转变——大数据 + 人工智能的机遇和挑战 [EB/OL]. [2025–01–02]. <https://www.shlab.org.cn/news/5443089>.
- 26 张智雄. 在人工智能时代贡献文献情报领域的智慧和方案 [J]. 农业图书情报学报, 2023, 35 (1): 5–8.
- 27 新一代人工智能发展规划 [EB/OL]. [2024–12–26]. https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm.
- 28 生成式人工智能服务管理暂行办法 [EB/OL]. [2024–12–26]. https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm.
- 29 “数据要素 ×” 三年行动计划 (2024–2026 年) [EB/OL]. [2024–12–26]. https://www.cac.gov.cn/2024-01/05/c_1706119078060945.htm.
- 30 刘辉. 卫生健康数据应用的机遇和挑战 [J]. 中国卫生, 2024 (7): 18–19.
- 31 上海发布卫生健康语料库暨首批医疗应用场景 MaaS 平台 [EB/OL]. [2024–12–26]. <https://www.shanghai.gov.cn/nw31406/20241128/ecf8db565a4944d39e2a72defbf a05e7.html>.