国内外数据重用研究述评*

林心悦[1,2] 杨建林[1,2]

[1]南京大学信息管理学院 南京 210023 [2]江苏省数据工程与知识服务重点实验室 南京 210023

摘要：［目的/意义］对国内外数据重用研究现状进行系统梳理，总结分析其呈现的特点与不足，并为未来数据重用相关研究提供借鉴。**［方法/过程］**运用文献调研法获取国内外数据重用相关研究文献，并基于内容分析法对其进行分类，总结当前数据重用研究所呈现的特点和存在的不足， 并提出后续研究建议。**［结果/结论］**现有数据重用研究在避免重复数据收集、提高数据使用效率和促进研究人员数据共享方面发挥了一定的作用，并且逐步关注了更广泛的用户群体、拓展和延伸了研究的学科和领域、关注的数据重用研究类型更加多元化；但总体研究方向较为狭窄、研究方法相对局限、研究数量相对稀少。未来数据重用研究领域，应进一步拓宽研究的用户群体、更关注社会经济发展过程中产生的新问题、更关注大数据时代对数据重用研究提出的新要求，进一步推动更有效和更可靠的数据重用研究，为科技创新、社会进步、国家发展贡献力量。

关键词：数据重用 数据管理 数据可重用性 重用困境

分类号：G350

1 引言

科学数据是国家科技创新发展和经济社会发展的重要基础性战略资源，也是科研活动的基础性资源,大部分的科研活动都是基于数据搜集和数据分析来开展的。1957 年，国际科学联合会理事会(International Council for Science，ICSU) 为了改善科学与技术数据的管理，从而提高数据的使用以此来促进科学发展，相继成立了国际数据组织世界数据中心（World Data Center，WDC)和科学技术数据委员会(Committee on Data for Science and Technology，CODATA)。近年来，我国持续注重科学技术的发展并着力投入资源，科研活动数量大幅提升，科研人员创新能力不断提高，科学数据也随之呈现出“爆发式”增长。尽管我国在科学数据管理与开放共享方面作了大量努力，但是存在诸多不足，例如科学数据共享的效率不高、范围有限，大量科学数据分散甚至流失、数据的价值没有得到最大程度的发挥[1]。这一局面是多种因素综合作用的结果，例如，缺少国家层面的法规保障，缺少专业的数据存储系统，科研人员对收集分析后的科学数据不再在意其潜在的价值、缺乏对其进行保存的意识不足以及专业的数据保存指导，等等。因此，为了提高科学数据的利用率，需要研究人员更深入地理解数据重用。

数据重用（Data Reuse）也被称为数据复用、数据再利用或二手数据使用。目前对数据重用的定义尚没有一个统一的定论，张潇月等根据定义侧重点的不

*本文系南京大学新时代文科卓越研究计划“中长期研究专项”项目（“数智赋能”背景下的情报学理论、方法与应用研究）的研究成果之一。

作者简介：林心悦，硕士研究生，Email：mf21140074@smail.nju.edu.cn。

同，将科研数据重用的定义分为4类:①“意图”派:注重从词语含义方面界定，突出原始目的以外的使用意图;②“情境”派:着重列举数据重用的具体情境;③“内容”派:对所重用数据的具体表现形式与呈现要求进行阐述，着重说明重用数据后形成的新的科研产出;④“流程”派:注重将数据开放共享与重用视为完整流程,强调数据重用过程必须可追溯[2]。重用科学数据，可以有效的避免重复收集数据，节省科研项目成本，节约科研人员的时间、精力。

部分学者曾对数据重用研究进行过述评，主要集中于数据重用困境解决措施，缺少全面的、系统的分析。为此，本文针对国内外数据重用研究进行分析，包括学者们关注的主题、用户群体、研究方法等，揭示研究呈现的新特点，述评国内外数据重用研究的特点与不足，为未来数据重用相关研究提供借鉴。

2 研究方法

2.1 分析样本

笔者采用中国知网、Web of Science平台作为数据来源，中文文献主要通过“数据重用”、“数据复用”、“数据二次使用”、“数据重复使用”作为主题词进行检索，选择“图书情报与数字图书馆”学科，英文文献采用Title=(data reuse) or (data reusing) or (data re-use) or (dataset reuse) or (secondary data reuse) or (data reusability) 进行检索，选择研究方向为“Information Science Library Science”，检索时间截至2022年3月14日，得到中文文献56篇，英文文献1944篇。通过人工判读，并对相关文献的引文和参考文献进行追踪，辅以Google Scholar，进行滚雪球式的追踪来补全因为检索词的不全面造成的漏检，最终获得98篇文献，形成本研究的分析样本。文献类型分布详见图1。从图1可以看出，实证调查类论文占比高达57%，是目前数据重用研究领域最主要的论文发表形式。

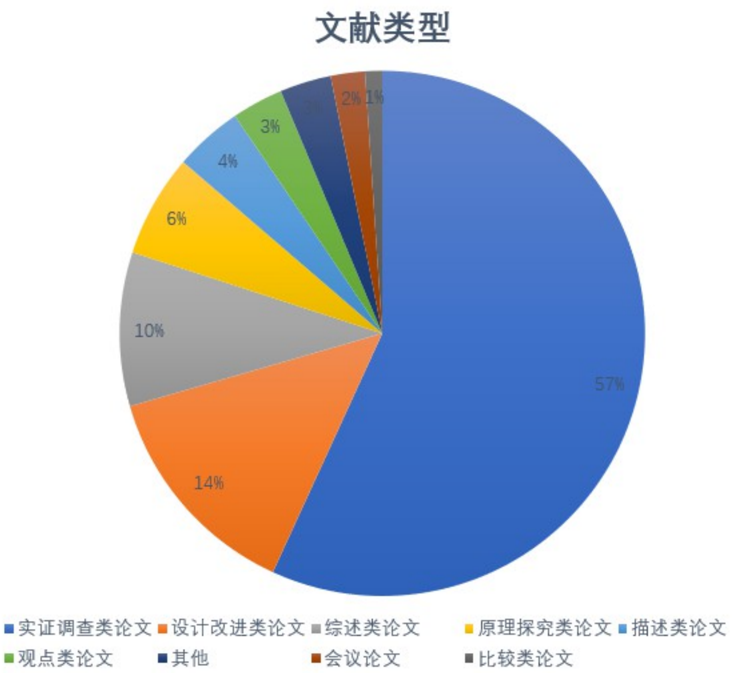

图 1 文献类型分布图

2.2 编码分析

本文采用内容分析的方法，对样本文献进行编码，从而识别数据重用研究的基本特点，梳理和总结国内外数据重用的研究进展与存在的问题。编码框架包括研究主题、研究对象、研究方法、使用或借鉴的理论模型及研究发现或结果，如表 1 所示。

表 1 文献编码框架

类目	说明/编码
研究主题	数据重用研究的不同方向，包括数据重用行为研究（包括群体对数据重用的态度看法、重用意愿以及他们重用数据的行为特点）、数据重用产生的利弊研究（包括数据重用产生的好、坏影响）、数据重用基础研究（包括对数据重用的概念、定义、过程、分类的研究）、数据重用影响因素研究（包括外部环境的影响因素，如政策法规、基础设施建设等；和内部影响因素，即重用者自身的因素）、数据可重用性评估研究（包括数据可获取性、数据的质量评估、数据可理解性、可信任性等判据）、综合性数据重用研究(仅笼统论述数据重用这一主题而难以从文中归纳出具体的研究类型)等
研究对象	研究的对象或样本涉及的群体，如社会科学研究人员、考古人员等；没有通过职业、年龄、身份等维度进行区分的研究对象，本文统称为综合人群
研究方法	研究所采用的方法，分为数据收集方法(如访谈法)和数据分析

	方法(如结构方程模型)
理论模型	用于支撑研究所使用或借鉴的理论或模型，如计划行为理论模型（TPB）、用户满意理论
研究结果/发现	主要研究结果和发现，以及研究呈现的创新之处

3 数据重用研究的特点

如图 2 所示，“数据重用行为研究”占比 37%，“数据重用影响因素研究”占比 22%，“数据重用产生的利弊研究”占比 12%，“数据重用基础性研究”占比 18%。这几个主题是数据重用研究的主要领域，总占比达到 89%。研究对象主要是科研人员，涉及主要领域为社会科学、STEM 学科、生物医学，此外教育学与教师、食品科学与营养学、天体物理、考古学等多个学科领域也得到了关注。

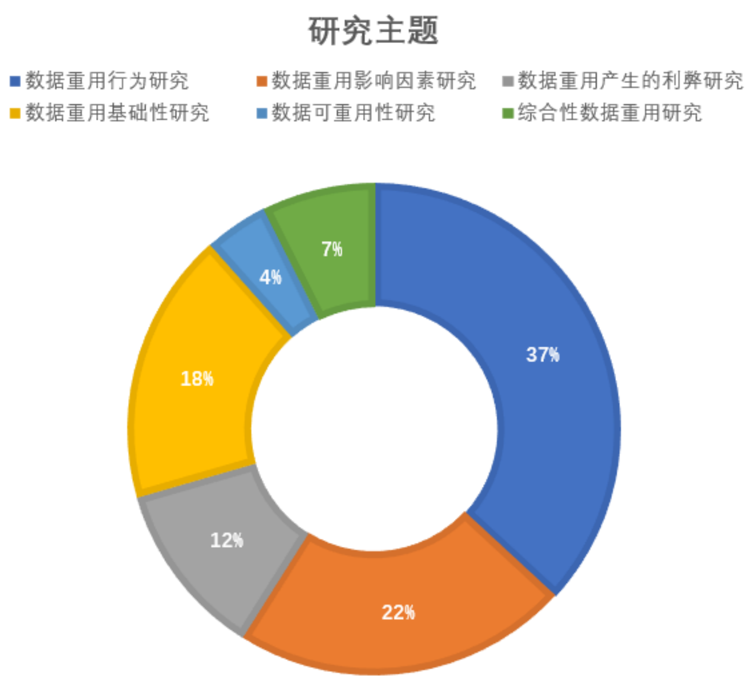

图 2 研究主题分布

表 2 展示了研究论文采用的数据收集与分析方法的分布。在 61 篇研究论文中，学者们多采用问卷调查（31.15%）、访谈法（18.03%）、检索（19.67%）、二手数据（8.20%）作为数据收集方法；结构方程模型（18.03%）和统计方法（9.84%）是问卷调查法常用的分析方法；针对通过访谈获得的数据，学者们多采用编码分析（16.39%）来进行分析；通过检索获得的数据，学者们多使用内容分析、统计分析的方法进行研究。

表 2 研究论文采用的数据收集、分析方法分布

数据收集方法	**数据分析方法**

方法	文献数量（百分比）	方法	文献数量（百分比）
二手数据	5（8.20%）	回归分析	3（4.92%）
		引文分析	1（1.64%）
		统计分析	1（1.64%）
访谈	11（18.03%）	解释性研究	1（1.64%）
		编码	10（16.39%）
检索	12（19.67%）	内容分析/编码	4（6.56%）
		统计分析	4（6.56%）
		回归分析	1（1.64%）
		引文分析	1（1.64%）
		混合方法	2（3.28%）
问卷	19（31.15%）	结构方程模型	11（18.03%）
		编码	1（1.64%）
		统计分析	6（9.84%）
		多元回归分析	1（1.64%）
小组讨论	2（3.28%）	统计分析	1（1.64%）
		编码	1（1.64%）
网络调查	2（3.28%）	案例研究	1（1.64%）
		文献分析	1（1.64%）
文本挖掘	1（1.64%）	统计分析/文献计量法	1（1.64%）
混合方法	5（8.20%）		
建模	4（6.56%）		

为开展研究，学者们使用、借鉴了多种理论模型，本文对此进行了统计。纳入“理论、模型的使用与借鉴”的标准为：①文献中详细介绍了相关理论、模型②研究设计基于该理论、模型，如基于某种理论提出了研究假设。对于仅引用而不在其基础上开展研究设计的不列入本文的统计范围[3]。表 3 中理论、模型的术语名称均来自于被分析的文献。

如表 3 所示，数据重用研究使用或借鉴最多的理论、模型是计划行为理论（TPB）、理性行为理论（TRA）。

表 3 数据重用研究使用或借鉴的理论、模型及频次

理论、模型	**使用或借鉴频次**
计划行为理论（TPB）	13
技术接受模型（TAM）	7
用户满意理论	2
继续使用(持续使用)意愿理论	2
Walsh and Downe model 、引文动机理论（Garfield，1979）、制度理论、激励理论/动机理论、使用统	1

一理论（UTAUT）、项目反应理论、Rasch 模型、生态系统模型、数据监护生命周期理论、MOA 理论、LDA 主题模型、信息使用环境理论、期望确认理论、感知风险理论、the Community Capability Model、自我决定理论、技术采纳与利用整合理论、理性行为理论（TRA）、IS 成功模型	

归纳起来，国内外数据重用研究的特点主要表现为：

（1）关注的用户群体广泛。数据重用研究主要涉及的群体是研究人员，考古学家、教师、文化中心与博物馆的工作人员等并非大量重用数据的群体也受到了关注。而研究人员所在的领域中，除了具有悠久传统共享和重用数据历史的社会科学、生物医药、STEM 学科以外，学者们不断拓展了研究的用户群体的所属学科，诸如营养学、食品科学、农业化学、天体物理学等。不过，尽管近年来数据重用的研究越来越多，但大多数研究都是针对特定学科领域中的数据重用实践研究，缺少对跨不同学科数据重用的研究[4]。

（2）数据重用行为研究受到最广泛的关注。数据重用行为研究又可以细分为数据重用行为意愿及影响因素、数据重用行为困境、数据重用行为促进策略等多个更为详细的研究领域。国内数据重用相关研究几乎都为行为研究，国外研究主题更为广泛一些，但行为研究仍然占据最大的部分。

（3）数据重用研究整体数量偏少，国内研究相比国外相对不足。仅从检索结果就可以看出，检索得到的数据重用相关文献中，国内文献为 23 篇，国外文献为 75 篇，且国内数据重用研究主题主要集中在行为研究，相比国外，研究范围相对狭窄，数据重用的定义、流程、框架等都鲜有研究。

（4）数据重用相关研究往往与数据共享有着紧密的联系。数据共享是数据重用的前提，数据重用是数据共享的目的，因此两者的研究往往相互涉及。从数据生命周期来看，数据共享与数据重用分别位于数据生命周期的不同阶段，数据共享涉及数据的处理、保存环节，是后续数据重用过程中数据检索获取、评估使用的前提。多项研究表明数据共享是影响数据重用的关键因素，而数据重用经验又会显著影响科研人员对数据共享的感知和共享规范。

（5）部分数据重用的相关研究，尽管得到了较多的关注和讨论，但由于问题的复杂性，目前仍没有一致的结论。如数据的可复用性评估作为数据重用流程中的核心阶段，从整体视角理解基于数据复用者感知的数据可复用性是一个非常复杂的问题，尽管有多位学者围绕数据本身(如数据可获得性、数据质量、数据存储库稳定性与安全性)、数据的生产者和使用者(可靠性等)、数据情境信息(如相关性)等多个维度进行了数据的可复用性的探讨，但至今仍没有得到其确定的内涵。

4 不同领域的研究进展

限于篇幅，本文选择几个重点领域，包括数据重用行为研究；数据重用概念、定义、框架研究；数据可重用性的评估与分析；数据重用带来的利弊研究，

阐述数据重用研究的进展。

4.1 数据重用行为研究

国内数据重用相关研究几乎都聚焦于这一领域，国外数据重用相关研究也有相当一部分关注于此，因此数据重用行为研究是数据重用相关研究中占比最大的一项。该领域着重探究研究人员对数据重用的态度、意愿、行为实践及相关影响因素。国内学者针对不同用户群体、不同学科的数据重用行为进行了研究，主要包括生物医学研究人员、社会科学研究人员等；外国学者进行了更为广泛的用户群体研究，主要包括 STEM（科学（Science），技术（Technology），工程（Engineering），数学（Mathematics））学科研究人员、生物医学研究人员、社会科学研究人员等。其他学科研究人员（天体物理学、动物学、食品科学、营养学和农业化学、教育学）和教师、考古学家、健康科学家的数据重用行为也受到了关注。此外，还有学者针对新手数据重用行为[5]和具有丰富数据重用经验的研究人员[6]进行了研究。

学者们针对数据重用行为意愿及影响因素、数据重用行为困境、数据重用行为促进策略等进行了广泛的探讨。

数据重用行为困境主要来源于法律法规、技术可行、认知接受三个维度。研究表明，当前的法律、政策并不能充分应对数据快速增长带来的数据重用需求的挑战。Helena Ursic等指出欧洲知识产权会限制数据重用者充分利用数据集、《数据保留法》中的数据本地化阻碍了国际数据传输，限制了全球范围内的数据交换和重用[7]。没有关于公开发布数据集以供重复使用的指导方针。例如，敏感数据是否以及如何被重复使用尚不明确[8]。Kathrin Dentler等指出尽管数据重用对病人乃至对社会有着巨大的好处，但是病人的数据涉及到隐私问题，需要得到保护。欧盟数据保护指令中指出：除非获得数据主体的同意或法律授权，否则个人数据不应被披露、提供或用于指定目的以外的其他目的的原则称为使用限制原则。此外，数据本身也是造成数据重用困境的主要原因，如数据质量、数据访问、数据可移植性等因素。Ayoung Yoon研究发现由于数据访问困难，一些参与者放弃了数据重用[9]；James W. McAllister指出经常缺少数据来源环境的相关信息，无法确保数据的准确性或无法理解数据，因此导致重复使用历史经验数据的困难和局限性。行为往往是态度的产物，态度反过来又受个人经验的影响。研究表明，科研人员对于数据重用的态度整体是积极的，但仍有许多担心。许多研究者不愿意与他人分享自己收集的数据，因为他们认为这是一种宝贵的竞争优势。数据重用被认为是减少数据收集费用和缩短研究过程的一种机制，对于时间和资源有限的研究人员来说，数据重用是一种可行且节约的选择，但是数据首次收集的意图和二次使用的意图往往并不匹配，这种不匹配可能需要额外的时间来整合数据，抵消了重用的好处。魏银珍等研究发现科研人员在数据重用过程中最为担心的是重用数据可能会带来的侵犯版权行为、对数据理解不够透彻、研究成果

发布受阻等问题[10]。一些科学家可能认为他们的数据对其他人没有价值仅仅是因为他们不知道其他人可以用它们做什么。如果数据被更广泛地共享，更多的科学家可能认识到他们的数据被意外使用的可能性，那么他们就更有可能进行共享。

由于数据重用具有诸多潜在的好处，学者们对促进数据重用的策略进行了研究。李佳潞等研究发现数据素养对科研人员的数据重用行为以及数据可重用性都具有显著的影响。较高的数据素养能够使科研人员感知到更高的数据仓储可获取性和科学数据的可重用性,同时具有较高数据素养的科研人员更有可能执行科学数据重用的实际行为。因此，广泛开展科研人员的数据素养教育、提高科研人员的数据素养有助于促进科研人员的科学数据重用行为[11]。张潇月等指出建优化开放科研数据基础设施环境、建立面向权益平衡的数据政策环境、科研支撑辅助机构提供的开放科研数据服务、关注科研人员的主观因素对科研数据重用行为的影响等都有利于促进数据重用[2]。数据共享为学术奖励体系带来了新的机遇。当科学家共享数据时，他们做出了重要的学术贡献，但目前还没有公认的方法来衡量和承认这一贡献。

4.2 数据重用概念、定义、框架、过程

数据重用至今没有官方的、确切的定义。广义的数据重用一词往往指数据的初始使用后的使用。狭义的理解上，数据是由一个人为特定的研究项目收集的，第一个用途是由该个人提出特定的研究问题。当该数据集被提交到存储库、由其他人检索并部署到另一个项目时，通常会将其视为重用。

一些学者对数据重用的类型做了区分。Bart Custers 等将数据重用区分为三类：（1）在同一上下文中为同一目的多次使用数据进行数据回收（2）数据重新调整用途—将数据用于与最初收集目的不同的目的，但仍处于与原始目的相同的环境中（3）在最初收集数据之外的另一个上下文中使用数据进行数据再上下文化[12]。

Xiaoguang Wang等分析了数据重用的过程。数据重用是一个动态的过程。初始阶段受数据需求的刺激，涉及是否重用现有数据的决策。第二阶段是探索和收集。在这一阶段，研究者需要以各种方式从各种来源发现、获取、理解和选择所需的数据，其中面向对象是数据实体和上下文。如果找到了足够的相关、有效和可用的数据，则会增强或放弃确定。当做出有利决策时，在选择和获得最终数据之前，研究者会在发现和获取数据以及理解和选择数据、纠正对不适当数据的选择或搜索额外数据以获得最佳拟合之间进行转换。数据选择后，收集的数据与研究目的相匹配，并开始二次处理，即重新调整用途。该阶段适应研究数据，采用多种数据处理操作，对二次处理数据进行研究[13]。

4.3 数据可重用性的评估与分析

Meloda 是 Alberto Abella 等人开发的一种评估数据可重用性程度的指标，该指标诞生于 2011 年[14]。Meloda 1.0 至 2.5 版本考虑了三个维度：数据集的技

术标准、访问数据的机制、数据的法律许可。随着在西班牙国家开放数据□户的数据集中应用，有人指出，有必要包括第四个维度，该维度将考虑要发布的数据模型，反映数据结构对处理信息（机器可处理）的重要性。改进后的 Meloda 3 包括四个维度：数据集的技术标准、访问数据的机制、数据的法律许可、数据模型 Meloda 4 则又增加了数据的地理信息和更新频率两个维度。Meloda 4 在五年的使用过程中，暴露了一些局限性。为了更深入地了解这一主题，一个国际专家小组就指标的两个方面进行了调查。第一个方面是，为了确定已发布数据集的可重用性，还应该考虑哪些其他因素。第二个方面是内部结构（即度量的每个维度的级别），它们是否应该增加、合并、删除或分割。最终小组考虑了两个新的维度：传播和声誉，并提出了新的内部结构、确定了每个维度的级别，得到了最新版本的 Meloda 5[14]。

表 4 Meloda 5

一级指标	二级指标
合法许可（最高6分）	1. 私用 2. 非商业性再利用 3. 商业再利用或无限制
获取信息（最高6分）	1. 对数据集的Web访问或唯一URL参数 2. Web访问对单个数据具有唯一的参数 3. API或查询语言
技术标准（最高6分）	1. 封闭式标准可重复使用和开放式不可重复使用 2. 开放式标准可重复使用 3. 开放标准、独立元数据
标准化（最高10分）	1. 自己的数据模型标准化 2. 发布自己的特殊数据模型标准化（协调） 3. 地方标准化 4. 全球标准化
地理定位内容（最多6分）	1. 没有地理信息 2. 简单或复杂文本字段 3. 坐标或完整的地理信息
数据更新频率（最高15分）	1. 超过1个月 2. 1个月到1天不等 3. 1天到1小时不等 4. 1小时到1分钟不等 5. 几秒钟（更新周期小于1分钟）
传播（最高6分）	1. 沟通/传播不系统 2. 更新可用资源（即RSS提要）

	3. 主动传播/推送传播（信息自动且及时）
声誉（最高6分）	1. 关于声誉的数据没有来源 2. 关于用户意见的统计或报告 3. 数据源声誉的指标或排名

Jihyun Kim等探讨了地震工程研究人员评估同事的实验数据可重用性的方法[15]。EE研究人员在评估数据可重用性时主要考虑三个问题：（1）数据是否相关，（2）数据是否可以理解，（3）数据是否可信，而评估途径主要包括期刊文章、个人网络、文档以及与产生数据的同事的对话，他们往往会同时通过多种途径，因为每种途径都有不同的能力支持他们用来评估同事数据的可重用性。

4.4 数据重用带来的利弊

数据重用可以给数据生产者和数据重用者两者都带来好处。对于数据生产者来说，数据被重用可以提高相应论文的影响力。Heather A. Piwowar等研究得出：在考虑了影响引文率的其他因素后，数据重用仍带来了强大的引文效益，并且第三方数据重用的直接影响在研究人员发表了大量重复使用自己数据的论文之后持续了多年。对于数据重用者来说，可以减少不必要的重复实验，缩短研究周期，降低科研成本，加快研究进程。

当然，所有事物都具有一体两面性。数据重用带来诸多好处的同时，不可避免的也具有一些副作用。Stefan Collini 在《The Slow Professor》的前言中写到：“当代学术界真正的知识生产力的障碍之一是大多数学者发表的文章太多。”尽管中共中央、国务院印发了《深化新时代教育评价改革总体方案》，强调“不得将论文数、项目数、课题经费等科研量化指标与绩效工资分配、奖励挂钩”但发表期刊数仍然是科学领域最重要的绩效指标。Erik M. van Raaij 指出在多个出版物中使用同一数据集可能意味着自我剽窃和重复、冗余、重叠出版物。Wiley Blackwell 在 2007 年的一项调查中显示，冗余出版是 16 种科研不端行为中最严重、最常见的一种，其次是剽窃、重复提交、未披露作者利益冲突。数据重用本身不是问题，但对跨多个出版物使用相同数据的实际案例的分析表明，过度和不当的数据重用可能造成学术不端[16]。

5 结语

科研数据不仅是科研活动的直接产物，更是支撑国家科学研究及科技创新的战略性资源。科研数据的重用已经引发国际组织、政府部门和研究机构的高度关注。总结过往以引领未来是一项重要的工作。本文通过分析 98 篇数据重用相关研究论文，梳理了国内外数据重用研究领域的进展并总结其特点，可以看到，在过去的二十年中，数据重用研究在减少重复数据收集、提高数据使用效率和促进研究人员数据共享方面发挥了一定的作用，并且逐步关注了更广泛的用户群体，拓展和延伸了研究的学科和领域，此外，研究类型也更加多元化。多位学者针对数据重用研究的重点领域展开了综述，包括数据重用行为、数据重用影响因素以

及数据重用产生的利弊的相关研究。同时，阐述了当前数据重用研究的问题与不足，如研究方向较为狭窄、研究方法相对局限、研究数量相对稀少等。

为此，后续的数据重用研究需在以下方面寻求突破：

（1）国内数据重用研究方向要更注重多样化。以往的数据重用研究群体较为单一，往往聚焦于高校、学者、研究人员，研究方向则多为行为研究，要突破局限，将研究视野拓宽到更广泛的人群、更宏观的社会层面，挖掘数据重用研究的深度关注整体社会的发展，切实帮助不同类型的群体，使数据重用研究能够服务社会、服务大众。

（2）重视大数据为数据重用研究带来的机遇和挑战。大数据革命已经波及到了各个领域，大量数据的积累为实现数据的再利用，继而最大化数据的价值提供了重要的基础。然而，如何开发利用这些数据，服务于社会经济、科学技术的发展，成为了数据重用研究领域面临的巨大挑战。这些挑战一方面来源于有待提升的数据存储技术和基础设施建设，另一方面也来源于数据保护法的限制。所以不仅要关注数据重用活动的价值，还要关注数据的质量，关注数据重用过程中的技术、伦理、法律法规等问题，以规范化地重用数据。

（3）加强数据生态系统的理论联系。数据生态系统是一个协同进化的整体，其应用理论亦不是独立形成的，如数据开放、数据共享是数据重用的前提，数据管理、数据素养是数据重用的重要影响因素。深化数据生态系统体系中的内部融合，能够为数据重用的进一步发展提供更有力的支持。

总之，未来的数据重用研究领域，应进一步拓宽研究的用户群体，更关注科学技术发展过程中产生的新问题，重视大数据时代对数据重用研究提出的新要求，从而为科技创新、社会进步、国家发展贡献力量。本研究也存在一定的局限性，分析样本只包括 CNKI 和 Web of science 两个数据库中检索后人工识别得出的文献，且由于采用结构化搜索策略，尽管尽可能地穷尽了文献中涉及的检索词，但仍无法避免检索命中率偏少的问题，本文主要通过“滚雪球”的方法来弥补这一缺陷；另外，美国国家科学理事会（National Science Board, NSB）将数据一词定义为“任何信息……包括文本、数字、图像、视频或电影、音频、软件、算法、方程式、动画、模型、模拟等”，仅以“Data”为核心词进行检索，未考虑其他形式的信息在不同学科都可能是数据的表现形式。因而，研究结果可能存在偏差。后续将继续追踪数据重用研究的发展，弥补这些局限性，从而更全面地揭示国内外数据重用研究的特点，以更好地促进数据重用研究领域的发展。

参考文献

[1]邢文明, 洪程. 开放为常态,不开放为例外——解读《科学数据管理办法》中的科学数据共享与利用[J]. 图书馆论坛, 2019, 39(1):8.

[2]张潇月, 顾立平, 胡良霖. 国内外开放科研数据重用困境解决措施述评[J]. 图书馆, 2021, 000(003):80-89.

[3]李月琳, 张建伟, 王姗姗,等. 回望"十三五":国内信息行为研究的特点,不足与展望[J]. 信息资源管理学报, 2022, 12(1):14.

[4] Kim Y , Yoon A . Scientists' data reuse behaviors: A multilevel analysis[J]. Journal of the Association for Information Science & Technology, 2017, 68(12).
[5] Faniel I M , Kriesberg A , Yakel E . Data reuse and sensemaking among novice social scientists[J]. Proceedings of the American Society for Information Science and Technology, 2012, 49(1).
[6]Yoon, Ayoung. Data reusers' trust development[J]. Journal of the Association for Information Science & Technology, 2016.
[7] Ursic H , Custers B . Legal Barriers and Enablers to Big Data Reuse[J]. Lexxion Publisher, 2016(2).
[8] Saxena S . Drivers and barriers towards re-using open government data (OGD): a case study of open data initiative in Oman[J]. Foresight, 2018, 20(2):00-00.
[9]Yoon, Ayoung. Red flags in data: Learning from failed data reuse experiences[C]// 2016:126.
[10]魏银珍，邓仲华，杨改贞. 科研人员数据重用意愿的影响因素研究[J]. 图书馆理论与实践，2020(3):6.
[11]李佳潞. 科研人员数据重用行为影响因素及促进策略研究[D]. 东北师范大学，2019.
[12]Custers B, Uršič H. Big data and data reuse: a taxonomy of data reuse for balancing big data benefits and personal data protection[J]. International data privacy law, 2016, 6(1): 4-15.
[13]Wang X , Duan Q , Liang M . Understanding the process of data reuse: An extensive review[J]. Journal of the Association for Information Science and Technology, 2021, 72(4).
[14]Abella A, Ortiz-de-Urbina-Criado M, De-Pablos-Heredero C. Meloda 5: A metric to assess open data reusability[J]. El profesional de la información (EPI), 2019, 28(6).
[15] Faniel I M , Jacobsen T E . Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data[J]. Kluwer Academic Publishers, 2010, 19(3-4):355-375.
[16]Erik, M, van, et al. Déjà lu: On the limits of data reuse across multiple publications[J]. Journal of Purchasing & Supply Management, 2018.

作者贡献说明：

林心悦：论文撰写与修改；
杨建林：论文指导。

Review of the research on data reuse at home and abroad

Lin Xinyue[1，2] Yang Jianlin[1，2]
School of Information Management, Nanjing University, Nanjing 210023[1]
Jiangsu Key Laboratory of Data Engineering and Knowledge Service, Nanjing 210023[2]

Abstract: [Purpose/Significance] To systematically sort out the research status of data reuse at home and abroad, summarize and analyze its characteristics and shortcomings, and provide reference for future research on data reuse. **[Method/Process]** Use the literature survey method to obtain domestic and foreign research literature on data reuse, classify them based on content analysis, summarize the characteristics and shortcomings of current data reuse research, and put forward follow-up research suggestions. **[Result/Conclusion]** The existing data reuse research has played a certain role in avoiding duplication of data collection, improving the efficiency of data use and promoting data sharing among researchers, and gradually pays attention to a wider user group, expands and extends the research discipline However, the overall research direction is relatively narrow, the research methods are relatively limited, and the number of studies is relatively sparse. In the future, the research field of data reuse should further expand the research user groups, pay more attention to the new problems arising in the process of social and economic development, and pay more attention to the new requirements for data reuse research in the era of big data, so as to further promote more effective and reliable data reuse. Research and contribute to scientific and technological innovation, social progress and national development.

Key words: data reuse data management data reusability reuse dilemma