DOI:10.13530/j.cnki.jlis.190026

科研人员数据复用行为研究:系统综述与元综合*

孙玉伟　成　颖　谢　娟

摘　要　学界已就科学数据复用的价值达成共识,不过实践却不尽如人意。文章采用系统综述和元综合方法对科研人员数据复用行为研究进行梳理,分别使用EBL和CIS对纳入文献进行批判性评估和解释综合。研究发现:不同学科的可复用数据源、数据评估判据以及复用行为影响因素存在情境敏感性和领域独立性;可复用的数据源主要有学科知识库、数据文档等正式渠道以及人际关系和联系作者等非正式渠道;判据主要源于数据的获取平台、自身、产生情境、生产以及复用者等;科研人员数据复用意愿和行为的影响因素涉及个人、机构/技术以及学科/社会等。后继研究可对不同学科领域的数据复用行为分别进行建模,采用整体和多维的理论视角对数据复用的影响因素展开实证研究。图1。表4。参考文献82。

关键词　数据复用行为　数据获取来源　数据评估判据　系统综述　元综合

分类号　G252.0

A Review on the Data Reuse Behavior of Scholars: System Review and Meta Synthesis

SUN Yuwei, CHENG Ying & XIE Juan

ABSTRACT

There is an academic consensus on the value of scientific data reuse, but the practice is not satisfactory. Through combing the main research problems of scholars' data reuse, the paper tries to clarify the situation of the research and to find out the gaps in the current research, then to present the future research directions. System reviews and literature meta-synthesis method were used in this study, which strictly controls the inclusion of sample articles. Then, a critical evaluation of the included articles was performed using EBL Critical Assessment Checklist. Thirty-nine articles that score more than 75% are included. Subsequently, bibliometrics characteristics and characteristics of the content were extracted. Finally, a Critical Interpretive Synthesis (CIS) method was used to synthesize the concepts of these topics on relevance criteria and perceived reusability of data, as well as the factors affecting data reusing behavior.

A series of results are concluded as follows: the research of data reuse behavior for scholars mainly focus on reusable resource, relevance criteria, and the willingness and influencing factors of reusing behavior. The reusable data sources mainly include formal channels such as printed literature, discipline repositories

* 本文系国家社会科学基金项目"施引者引用意向与文献计量视角的学术论文被引影响因素研究"(编号:17BTQ014)的研究成果之一。(This article is an outcome of the project "A Study of the Factors Affecting Citation Counts from the Perspective of Scientists and Bibliometrics" (No. 17BTQ014) supported by National Social Science Foundation of China.)

通信作者:成颖,Email:chengy@nju.edu.cn,ORCID:0000-0002-0664-7206(Correspondence should be addressed to CHENG Ying, Email:chengy@nju.edu.cn, ORCID: 0000-0002-0664-7206)

and data documents, as well as informal channels such as interpersonal relations and contacting authors. The essence of data assessment is data reusability assessment, including the assessment of data quality, data trust, the quality of data document, etc. Data assessment involves criteria derived from data itself, data producers, data reusers, data context information and acquisition platforms, etc. Influencing factors that affect the data reuse behavior and researchers' willingness to data reuse involve personal factors, institutional/technical factors, discipline/social environment factors and other relevant factors.

In different disciplines, there are situational sensitivity and domain independence of reusable sources, data evaluation criteria, data reuse behavior and its influencing factors. Therefore, there is no one-size-fits-all solution to understand the data reuse behavior of all disciplines, and it is necessary to determine the needs of different stakeholders according to their disciplines and to model the process of data reuse accordingly. It is also important to clarify the relationships between reusability and data quality, trust and the quality of data documents, different theoretical perspectives and multi-dimensional research methods are necessary for the empirical research on the influence factors of scholars' data reuse.

The limitation of this study lies in that the literature search strategy does not take into account the retrieval terms of various Data forms in the definition of Data, and only using "Data" as the retrieval term inevitably leaves out some Data reuse studies. However, this study comprehensively reviewed the research of data reuse from the perspective of scholars and it provides references for future research and scientific data management practice of library and intelligence institutions.1 fig. 4 tabs. 82 refs.

KEY WORDS

Data reuse behavior. Data source. Data evaluation criteria. Systematic review. Meta synthesis.

0 引言

科研人员数据复用(Data Reuse),也译为"数据重用""数据再利用",指的是为了新的研究目的对数据的二次使用[1]。数据复用以数据共享为基础,继美国 NIH、NSF 以及欧盟提出数据共享政策之后[2-4],2018 年 1 月,中共中央全面深化改革领导小组审核通过《科学数据管理办法》,强调"加强和规范科学数据管理,要适应大数据发展趋势,积极推进科学数据资源开发利用和开放共享",这些政策对科研人员数据共享与复用实践起到引领作用,为数据复用创造了有利的条件。目前,数据复用在自然科学领域(如基因组学、地球科学、天文学、高能物理等)和社会科学领域(如政治学、社会学等)均有较长时间的历史,数据复用规范认可度较高,其数据复用基础设施也较为完善,如基因学领域的基因序列数据库 GenBank[5]、地球科学领域的地球数据观测网络 DataONE[6]、天文学领域的 SDSS[7]、跨校政治与社会研究联盟 ICPSR[8](定量数据)、家庭生活与工作经验 QualiBank[9](定性数据)等,而在某些学科如考古学[10]、计算机科学[11],数据复用规范或文化认可度比较低,还仅仅存在于人与人之间的交流之中。尽管不同学科对数据复用的认知还存在差异,但是数据复用给科学研究带来的价值已成为学界共识[12-15]。

近年来,为促进科研人员数据复用,相关学者主要从两个视角展开研究。其一是从数据管理者的视角,科学数据管理作为图书情报机构新兴的服务内容得到研究者的持续关注[16-17],数据管理基础设施要保证学术记录的完整性和真实性,保证数据以可访问和可理解的形式呈现给用户,以支持数据复用[18-19];为保证数据管理的有效性,图书情报机构开发了系列标准和

方案,如数据类型和格式标准(如 PRONOM1, Research Data Alliance2)、元数据方案(如 schema.org)以及数据监护需求调查模板(如 Data Curation Profiles3)等[20],其目的是通过对科学数据的持续监护实现数据的可持续重复使用。其二是从数据复用者的视角,研究科研人员数据复用行为,对数据复用者在数据复用过程中的数据需求、数据获取、数据处理、数据评估及使用,以及数据复用态度、意愿、行为影响因素等进行研究,代表性的学者有 Faniel[8,21-24]、Yoon[1,25-29]、Kim[30-31]、Zimmerman[32-34]、Curty[35-36]、Murillo[6]、Daniels[37]等。两种视角下的研究得出了大量有价值的洞见,用以辅助数据复用政策制定、基础设施建设等,尽管如此,科研人员数据复用实践仍然很不普遍[10-11],科研人员难以获取或者无法复用共享数据,呈现出对数据复用价值的高度认可与低水平数据复用实践之间的矛盾[38-39]。因此,非常有必要对科学数据复用行为研究进行系统梳理,从而厘清主要研究问题,探索该领域的研究不足并提出进一步的研究思路,具体问题包括:可复用的数据源有哪些,数据评估的判据是什么,数据复用的态度、意愿、行为及其影响因素是什么。

1 数据与方法

文章利用系统综述(Systematic Review)和文献元综合(Meta Synthesis)方法对科研人员数据复用研究进行数据提取和综合。文献元综合是一个对特定主题的定性和定量研究进行严格搜索、评估、综合和解释的研究方法,通常包含四个步骤[40]:①制定高度结构化的搜索策略,作为研究人员查找主题文章的依据;②确定纳入和排除标准(如数据范围、主题焦点等);③基于明确的批判性评估标准对文献进行评估;④采用合理的方法进行数据提取和综合[41-42]。文献元综合方法以系统综述方法为基础,但又不同于系统综述,主要表现在步骤③和步骤④的差异上,文献元综合有明确的批判性评估标准,并基于合理的方法进行综合;文献元综合也不同于元分析(Meta-Analysis),元分析使用统计学检验将多个定量研究结果综合起来[43],而元综合则可以将定性研究和定量研究的结果一同纳入研究范围进行解释综合[44]。

目前,文献元综合方法在图情领域主要应用于信息行为研究,Urquhart 和 Yeoman[44-45]分别探讨了元综合方法在信息查询行为研究中应用的可能性,并对女性的信息行为的不同研究主题进行比较综合,确定了性别变量是否应该作为信息行为研究的主要变量;Urquhart[46]通过分析不同的元综合方法发现,可以用现实主义的综合或批判性的解释综合来整合信息行为研究与信息素养研究,Catalano[40]则真正应用批判性解释综合法对有关研究生的信息素质和信息行为原始实证文献进行综合,提炼出研究生的信息查询模式。科研人员数据复用行为研究涉及多种研究方法,以基于访谈的定性研究和基于调查的定量研究为主,这与用户研究视角下的信息行为研究一脉相承,因此本文尝试利用元综合方法对科研人员数据复用行为进行提炼综合。

1.1 检索策略

(1)检索词选择:通过初步阅读数据复用相关文献,设定检索词的可能范围:"Data Reuse" OR "Data Reusing" OR "Data Re-use" OR "Dataset Reuse" OR "Secondary Data Use" OR "Data Reusability","数据复用"OR"数据重复使用"OR"数据二次使用"OR"数据重用"。

(2)检索源选择以及检索方式:选取 WoS 核心合集、Scopus 文摘数据库、Proquest 硕博士论文全文库、中文论文 CNKI 数据库,在每一个数据库中对相关文献的引文和参考文献进行追踪,并辅以 Google Scholar 和必应学术搜索以及个人学术网站,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。

1.2 纳入标准

确定纳入标准是进行系统综述和文献元综

合分析的必要条件,根据研究目标和研究问题,本文确定了如下纳入标准:①以科研人员为研究对象的数据复用研究,包含探索性和验证性的实证研究;②文献类型仅限于中英文期刊论文、会议论文、学位论文;③将同一作者的相似内容的研究合并,如 Yoon 和 Murillo 均是数据复用研究的主要作者,以博士论文为基础发表了相似会议论文和期刊论文,因学位论文研究更系统,保留学位论文,若期刊论文和会议论文是博士论文研究内容的延伸则保留;④EBL 批判性评估大于 75%。

具体做法如下:以 1.1 中阐述的检索策略检索 WoS 数据库,共获得文献 1 128 篇,通过浏览标题(Title),初步得到 35 篇,剔除思辨性研究 3 篇以及采用文献计量和内容分析方法开展的二手研究 3 篇,共获得数据复用的探索性与验证性研究文献 29 篇,对这 29 篇论文进行引证文献和参考文献追踪后共获得研究文献 45 篇;在 SCOPUS 数据库和 Proquest 学位论文数据库采用同样的方法分别得到 38 篇期刊论文和 7 篇学位论文,对其参考文献和引证文献追踪后分别得到 49 篇和 25 篇研究论文;对三个来源的文献合并去重后得到 56 篇期刊论文和 7 篇学位论文;后通过 Google Scholar 检索并浏览个人学术主页增补文献 1 篇;在中文数据库 CNKI 中未命中相关文献;因此最终进入 EBL 评估的文献有 64 篇,取其中 EBL 得分大于 75%的 39 篇为本研究纳入分析的文献(见图 1)。

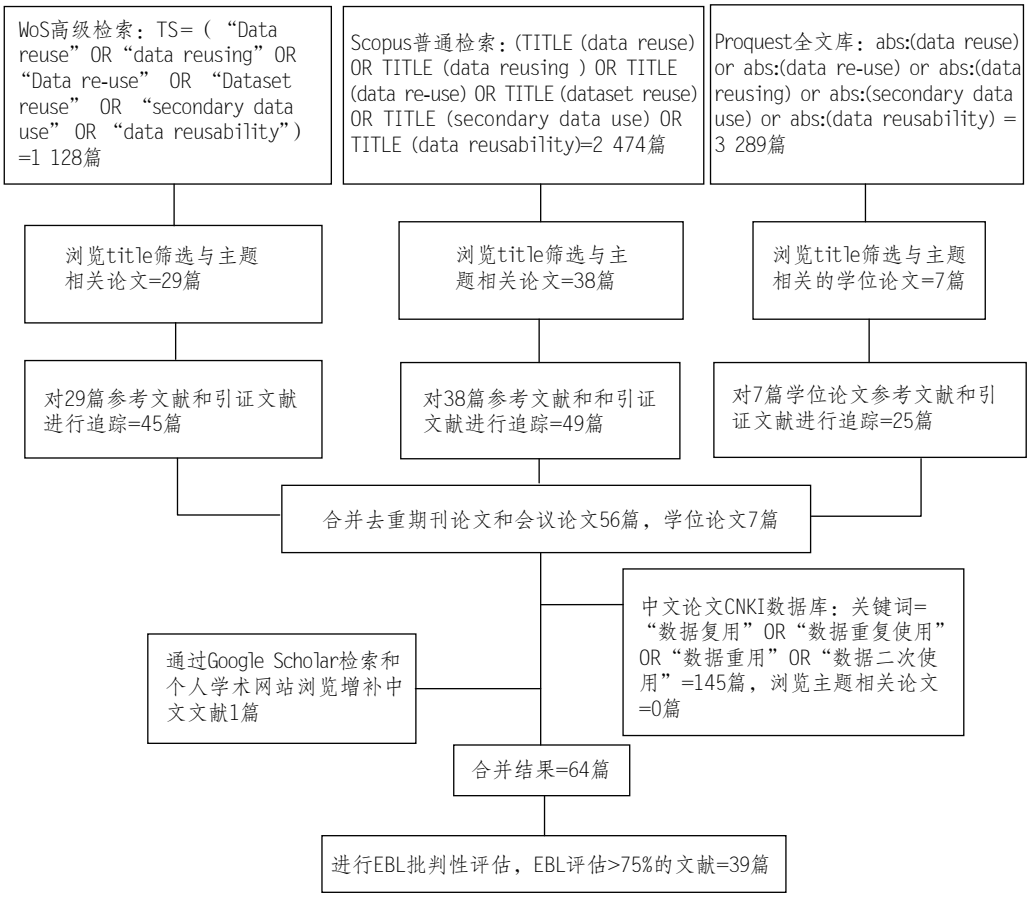

图 1 论文检索与评估流程

1.3 批判性评估(Critical Appraisal)

“批判性评估”起源于循证医学,用于评估研究的有效性[47]。对应于不同的研究领域已有一些评估检查工具,如健康或医学领域的 Crowe 批判性评估工具[48]、教育领域的 ReLIANT 文献评估读者指南[49]、图书情报领域的 EBL[50-51]等。本研究选择 EBL 对纳入文献进行批判性评估。Glynn 检查表包括四部分:①总体特征:样本的选择及纳入标准等;②数据收集:方法、量表有效性、是否消除偏见、包含证据等;③研究设计:研究方法是否适用、是否有表面效度、研究方法是否足够详细以利于复制、是否通过伦理审查等;④结论:是否有合理的结论,适用于指导实践等[50-51]。在使用此检查表对文献进行评估时,对应于检查表中的四个部分分别对照评估,总分数超过 75%说明研究是有效的,表 1 中列出了 75%以上的有效研究样例。

需要说明的是,检查表中的 Section A 部分

表 1 纳入 EBL 评估的文献集(样例)

作者及发文年代	来源	学科	样本量	研究方法	取样方法	数据收集工具	EBLIP 分(%)	简评	研究主题/研究目的
Daniels, 2014	University of Michigan	考古学、植物学	41	案例比较、半结构化访谈	立意抽样(最大变异抽样)、滚雪球抽样	访谈提纲	>75	访谈数据收集减少了偏见,稳健的研究设计	研究人员如何发现、评估、分析和使用数据
Stvilia et al, 2015	Journal of the Association for Information Science and Technology	物理学	12, 160/672	访谈、在线调查	立意抽样	访谈提纲、自建量表	>75	知情同意,访谈提纲未包含在出版物中	物理学家感知的数据质量
Sands et al, 2012	Proceedings of The Asist Annual Meeting	天文学	14	访谈	立意抽样	访谈提纲	>75	访谈提纲未包含在出版物中	天文学家发现、定位、检索外部数据的模式
Zimmerman, 2007	International Journal on Digital Libraries	生态学	13	半结构化访谈	立意抽样	访谈问题	>75	描述统计	数据复用的实践:数据发现、获取、确认过程
Faniel et al, 2016	Journal of the Association for Information Science and Technology	社会科学	249/1480	在线调查、文献计量	立意抽样	自建量表	>75	获得知情同意,低反应率,描述统计和推断统计	数据质量与满意度之间的关系

续表

作者及发文年代	来源	学科	样本量	研究方法	取样方法	数据收集工具	EBLIP分(%)	简评	研究主题/研究目的
Faniel et al,2013a	Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries	考古学	22	访谈	滚雪球和便利抽样	访谈提纲	>75	脚本编码经过偶然一致性的信度检验(scott's Pi 0.73)	数据复用的情境影响因素
Murillo,2016	University of North Carolina at Chapel Hill	地球科学、环境科学,生态学和生物学	16位科研人员、157条元数据记录	准实验出声思维、实验后调查、内容分析	随机抽样、分层抽样	计算机下载、出声思维指南、自建调查问题	>75	出声思维研究样本小,Section A <% 75,描述统计	数据可复用性评估、数据复用行为
Joo et al,2017	Aslib Journal of Information Management	健康科学	161	调查	随机抽样	自建量表	>75	样本量和反应率低,Section A <% 75,研究设计很完备,推断统计	数据复用行为及其影响因素
Kim& Yoon,2017	Journal of the Association for Information Science and Technology	STEM学科	1237/1528	调查	随机抽样	自建量表	>75	人群选择确定了纳入和排除标准,完备的研究设计,描述统计和推断统计	数据复用意图及其影响因素
Yoon& Kim,2017	Library & Information Science Research	社会科学	292/2193	在线调查	随机抽样	自建量表	>75	强健的研究设计,推断统计	社会科学家的数据复用行为

"样本大小是否足够大,可以得到足够精确的估计?""反应率是否足够大,可以得到足够精确的估计?"以及 Section B 部分"数据收集工具是否得到验证?"三条评估标准,仅仅针对调查研究中的随机抽样,其评判标准采用 Israel 2009 年建立的对应表[52](其中置信区间为 3%—10%,置信水平设置为 95%,P=0.5),而"响应率是否足够大,可以得到足够精确的估计?"的评判,采用在线调查的样本反应率应大于 30%,面对面调查的样本反应率大于 80%的标准[53]。对于质性研究(如访谈法),以上三个标准是不适用的,排除在计算之外。

1.4 数据抽取与综合

(1)数据抽取

从每一篇纳入研究范围的文献中抽取文献计量特征、研究设计、研究目标和结论等信息,并用EXCEL表格整理,如表1所示,其中包含了纳入研究范围的每一篇文献的作者、出版年代、文献类型、研究学科、数据收集方法和工具、研究样本、取样方法、研究主题和研究目标、EBL批判性评估分数及简评信息,简评是评判性评估过程中对评估列表中的重要结论进行说明,如表明推断统计或描述统计,有无样本选择偏见,样本反应率以及有效性计算的中间分数等。另外,作者所属地区或国家限于对第一作者单位的抽取。以上数据提取由两位作者进行独立评估,每项研究的细节都被提取出来,然后对比差异,最后通过讨论和协商解决,以确保数据提取的一致性。

(2)数据综合

在识别研究内容维度的基础上,对研究内容采用元综合中的"批判性解释综合"(Critical Interpretive Synthesis, CIS)[54-55]的方法对研究问题进行综合,CIS本质上是一种"解释"的探究模式,是通过回顾和综合的过程来发展新的概念和理论。它可以综合定性研究和定量研究的证据,比如利用定性数据解释定量研究存在的不一致性,从而解释现象[56]。在具体过程中,CIS通过对每一篇文献进行研读并记录笔记,提取研究的主要发现,科研人员复用研究的主要研究主题,涉及数据获取、数据评估、数据复用感知、意愿和行为等核心概念,对核心概念内的数据获取来源、数据评估判据、数据复用行为影响因素进行批判性解释综合。

2 结果

表1中包含了纳入EBL评估文献的作者、发表年代、研究方法、数据收集工具、取样方法、EBL评估得分以及评估过程中需要强调的内容(简评)等。39篇文献中,文献类型分布为:期刊论文22篇,会议论文9篇,博士论文7篇,研究报告1篇;文献的发文年代分布为:2000—2009年5篇,2010—2014年14篇,2015—2018年20篇;从文献的年代分布可见,该主题研究文献量呈上升趋势,尤其是近几年的论文占了总文献量的近一半,可见科研人员数据复用行为是一个新颖的研究主题。目前,相关研究主要由国外学者完成,我国除了台湾地区外,尚未见直接以科研人员"数据复用"为题的研究。

39项研究中,EBL评估都超过75%,表明纳入研究在研究设计、实施以及结果呈现等方面都较规范。具体到研究方法,除了少数研究利用观察[11]、准实验[6]、案例场景法[5,57]、内容分析[6]等方法外,更多地使用了访谈(25篇)与问卷调查(17篇),访谈以半结构化访谈为主,样本量在6—60之间;问卷调查则以在线方式为主,样本量在151—2 277之间;很多研究综合使用了两种以上的研究方法,如Faniel[21-24]、Murillo[6]等的工作。在39项研究中,有15项明确了研究的理论基础或框架,如数据复用者的信任理论基础来源于社会学、社会心理学、信息和信息系统相关理论[1],生态学的科研人员数据复用实践研究来源于测量理论和实践社区理论[34],基因学科研人员数据质量评估来源于活动理论[57],社会科学家数据复用的意图和行为来源于信息系统研究的集成技术接收与使用理论模型(UTAUT)[35],而相对集中的理论基础来源是社会心理学的理性行为理论[37]和计划行为理论[29,31,58-59],利用这两种理论主要研究了数据复用意图和行为。

2.1 可复用的数据源

纳入研究中有14项涉及可复用的数据源问题(见表2)。综合相关研究可以发现,复用的数据主要来源于纸质文献(论文、专著、报告、目录、统计资料等)、学科知识库/在线数据库/数字仓储、资料说明书(数据文档)等正式渠道以及人际关系和联系作者等非正式渠道。

在正式渠道中,大部分研究都会使用期刊论

表 2 科研人员数据获取来源

数据获取来源	
正式渠道	纸质文献(期刊论文、专著、报告、目录、统计资料、已发表的记录)
	学科知识库(数据存储库)
	在线数据库(综合数据库)
	数字仓储
	资料说明书(数据文档)
	搜索引擎
	学会或调查机构等的推广活动(调查数据)
	机构记录
	政府或学术网站
	博物馆(野外笔记、照片、领域笔记、挖掘报告、现场草图、未出版的手稿、博物馆入藏登记簿)
	天文台(观察结果、调查数据)
非正式渠道	人际关系(同事、导师等)
	数据作者(提供者)

文、著作、报告等纸质文献作为数据渠道或来源,既可以根据这些文献提供的线索获取数据,也可以直接从中找到相关数据,比如原始论文对数据集使用的说明可指导数据复用者理解数据的收集方式、内容限制等[23];学科知识库是查找数据的重要渠道,比如生态学领域的生物控制数据库、林业数据库、年轮数据库、气候数据库[32],基因领域研究中的 GenBank[5],天文学领域的 SDSS[7];资料说明书(数据文档)(Documentation)是又一重要渠道,它是对数据的收集、测试、分析与处理过程的记录[23],记录了数据产生的详细信息,是科研人员获取数据时的有效补充。

在非正式渠道中,大部分都提到了联系作者和人际关系渠道,比如环境科学 40%的数据来自数据提供者[60],嵌入式网络传感中心(CENS)的地球和环境科学研究人员偶尔从注册中心和个人获取数据,数据获取只发生在正常的人际交往中[11],生态学家求助于个人关系来找到相似的数据[32],天文学家数据获取的渠道之一是联系作者[7],野外考古学家主要依靠同事和博物馆馆长来定位数据和相关的背景[21],年轻的定量社会学研究人员主要依靠有数据复用经验的教师顾问,因为教师顾问能够找到相关数据并理解数据的局限性[8]。并且由于不同学科科学研究数据特征的差异,使得不同学科的学者通过人际关系询问的侧重点不同,比如:地震学者向同事问资料收集仪器、型号、材质及资料的信效度等[23],社会科学学者则会询问质化研究过程中的情境信息[28]。

一般来说,大部分学科的学者都会选择多种数据来源渠道收集数据,同时使用正式渠道和非正式渠道;不同学科的数据来源偏好也有差异,自然科学学者偏好从文献或学科机构库中获取数据,社会科学学者偏好从人际关系中获取数据。

2.2 科研人员数据评估的判据

在科研人员数据复用过程中,数据评估是一项复杂且重要的中间环节,在 39 项研究中,有 22 项涉及数据评估及相关判据的研究,利用系统综述主要发现以下研究主题:数据的可复用性评估[6,23]、数据信任评估[1,21,23,27]、数据质量评估[5,20,22,25,57]、数据文档的质量评估[61]等(见表 3)。对定性和定量研究得出的数据复用评估判据分研究主题进行批判性解释综合[54-55],通过比较相同证据在同一研究主题的不同研究中的重要性排序,对数据评估判据进行解释。最后,综合不同研究主题的证据发现,数据的可复用性评估是数据信任评估、数据质量评估、数据文档质量评估的上位概念,数据评估的本质是数据可复用性评估。

(1)数据可复用性评估

科研人员判断数据是否可复用涉及数据复用过程的主要环节,首先数据是否可获取是数据复用的前提条件,获取到数据后对数据进行评估,评估会涉及数据的质量评估[5,20,57]、相关

表 3 科研人员数据复用评估判据

研究主题	评估判据	学科
数据可复用评估	数据描述/摘要信息、数据属性列表、研究方法信息	地球科学、环境科学,生态学和生物学
	数据相关性、可理解性(元数据情境信息的完整性)、可信任性	地球科学
	数据文档(数据说明书)、数据适用性、数据生产者的可信任性/信誉、数据质量、复用准备	社会科学
	数据生产者的研究方法、数据生产者如何进行研究的情境信息	定量社会学、考古学
	领域知识、个人数据收集的经验、对领域研究趋势的熟悉程度	生态学
数据信任评估	数据的质量、合并的来源、元数据和数据文档的充分性、数据处理方式等	天文学
	数据收集过程、采样方法、采集程序、测量和编码、量表的信度、效度、数据的研究方法设计、数据文档(描述选择和调整数据收集工具的文档)	社会科学、考古学、地球科学、环境科学,生态学和生物学
	数据工具细节	地震工程、天文学
	数据来源的声誉和权威性、数据生产者及数据存储库的声誉、与数据生产者是否熟悉	基因学、考古学、天文学、生态学、社会科学
	数据来源和数据文件的专业性、数据相关性和可用性、数据可靠性和有效性、	社会科学、地震工程
数据质量评估	有用性、相关性、精确性、数据可获得性、安全性等	基因学
	准确性、简便性、信息化、可访问性	物理学
	相关性、样本的代表性、测量的信度和效度、可用性(可获得性、易操作)、数据文档的可理解性以数据生产者的努力程度和承诺、数据生产者的伦理和意图	定量研究社会学
	实地或实验室的经验、领域知识	生态学
	数据作者的声誉和学术地位、在实地工作中创造出来的描述措辞和结构,以及关于存放数据的仓库的信息	考古学
数据文档的质量评估	易于使用、充分性以及准确性,数据生产者的动机和存档数据的能力、数据使用者的吸收能力、改进文档质量的中介(如数据档案)、数据对隐性知识的解释能力	社会科学

性判断[6,23]、数据是否可理解[23]、是否可信任[23]等判据。相关性指的是数据回答研究问题的程度,可理解性指的是数据的预期含义是否容易被感知,可信度是研究人员可以在多大程度上信任他人创造的数据。数据的相关性判断主要依靠数据产生的情境信息[6,8,23]、数据说明书(数据文档)[35]等,数据的情境信息包含数据描述/摘要信息、数据属性列表、研究方法信息等[6,8],此外,数据复用者的领域知识、个人经验以及对领域研究趋势的熟悉程

度[32-33]也是数据是否可复用的判据。

值得一提的是,在社会科学数据复用评估判据中,数据是否可复用与数据复用满意度是有区别的,相同判据在不同的研究主题中会处于不同的位置,当考虑数据是否可复用时,相关性、可信性、可理解性、可操作性是首要考虑要素,完整性、可获得性、数据生产者的声誉在其次。当考虑数据复用满意度时,可获得性、完整性、文档质量、可信性和可操作性是首要考虑要素,相关性、期刊排名以及数据生产者的声誉则与数据复用的满意度相关度很低[22]。

(2)数据质量评估

数据质量可以定义为满足数据使用活动的程度[20],好的数据质量就是满足用户的期望和注释活动需求的数据[57]。在数据质量判据中,大部分相关研究把准确性(Accuracy)作为数据质量评估的第一判据[5,20,57]。如 Huang 将准确性、无偏误(Unbiased)、可信度(Believability)、可追踪度(Traceability)都归为准确性维度下,认为准确性和可访问性是最重要的数据质量维度[5];Stvilia 等人利用主成分分析法识别出影响数据质量的 13 个指标中,正确性位居第一位[20];数据可获得性是数据质量评估的又一重要判据[5,20,22,25],在 Huang 对基因学和 Stvilia 等人对物理学的数据质量判据研究中位居第二位;相关性也是数据质量评估中提到较多的判据[5,20,22,25,57],但其重要性排序均在第二位以后(基因学领域排在第 4 位[5,57]、物理学领域排在第 9 位[20]),并且在社会科学领域,相关性作为数据质量的一个维度,其与数据复用满意度之间的关系未得到验证[22]。除以上判据以外,特定学科还有一些个性化的判据,如定量社会科学的判据还有样本的代表性、测量的信度和效度、数据文档的可理解性、数据生产者的努力程度和承诺以及数据生产者的伦理和意图[25];生态学的还有实地或实验室的经验以及领域知识[33];考古学的还有存放数据的仓库的信息[8,21]。

(3)数据可信任性评估

现有的研究表明,对现有数据缺乏信任以及对数据复用的感知规范没有被发现是数据复用的主要障碍[36],研究人员需要在使用共享数据之前建立信任和可靠性[23,32],关于数据信任的判据,数据本身的特征、数据的质量[62]、研究过程信息如数据的采样方法、采集程序、测量和编码、量表的信度效度等[8,23,24]、数据的研究方法设计[6,24,61]、数据工具细节[23,62]都会影响数据信任;数据的情境信息如数据是如何产生和处理的[28]、元数据的充分性[62]、数据文档的完备性[62]也会影响数据信任。一般而言,数据产生的背景信息越丰富、越准确,可信度越高。另外,数据来源的声誉和权威性[21,27,33,37,57,62]、数据存储库的声誉[21]、与数据提供者是否熟悉[28]均是数据可信任性的判据。以上数据信任的判据是将数据复用作为一个整体研究对象考量的,如果将数据复用的过程分阶段考量,研究人员发现,在社会科学整个数据复用的过程中,数据信任的判据呈现动态性,在数据发现和选择阶段,其信任判据是数据来源和数据文件的专业性、数据的相关性和可用性;在数据获取、审查和理解阶段,数据信任判据是可靠性(信度)和有效性(效度);在问题解决阶段,其判据是数据生产者及数据存储库等。最初的信任在数据复用的早期阶段发挥了重要作用,引导用户进入下一个阶段,中间的信任(临时信任)在数据审查过程中有可能降低也有可能提升,最后,问题解决阶段的信任程度会决定数据使用者是否对数据进行重新确认[1,27]。

(4)数据文档质量评估

社会科学数据文档是数据生产者在数据产生过程中记录的用以帮助数据复用者进行评估数据的详细知识,有多种类型,如数据字典、数据收集工具、基于数据的出版物、用户指南、统计手册等。数据文档有时也称作元数据,是一种用来进行资源发现和数据二次分析的元数据,比如数据文档倡议(Data Documentation Initiative,DDI)记录的是结构化元数据,而访谈或调查问题的文本则是非结构化元数据。研究表明,数据复用需要深入理解数据创建时的情境信息[24,32-33,63],数

据文档可以很好地传递情境信息,对研究人员数据复用体验的满意度具有正向影响[22,61]。在社会科学领域,Niu 将科研人员感知的数据文档质量评估要素划分为易于使用(Ease-of-use)、充分性(Sufficiency)以及准确性(Accuracy),每个要素又划分为几个子要素,建立了数据文档评估模型并对模型进行验证,结果表明感知的文档质量受数据生产者的动机和存档数据的能力、数据使用者的吸收能力、改进文档质量的中介(如数据档案)以及数据对隐性知识的解释能力的影响[61]。

在上述判据综合的基础上,将数据评估判据进行整合(见表 3),其中数据的情境(Context)信息、数据质量信息、数据信任信息是不同学科数据评估的共同判据。数据信任、数据质量均作为数据可复用性判据,两者存在相互影响关系。数据质量是数据信任的判据,数据的可信任性也是数据质量的属性,如 Stvilia 等发现,在 13 种影响数据质量的指标中,可靠性(Reliability)与可验证性(Veriability)排序分列第二、三位,这两个指标与数据的可信任性有关[20]。情境信息提供数据生产者如何进行科学研究的相关信息[8],是对数据生成的相关环境的描述的集合,包括数据产生的物理环境、数据获取的技术和社会环境等[23],数据的情境信息贯穿于数据可复用性评估、数据信任评估、数据质量评估中。在评估过程中,研究人员需要尽可能多的情境信息判断数据的相关性,与情境因素相关的数据文档可获得性和完备性[24,26,33,37,63]、元数据/数据描述信息[23,37,63-64]等都是数据可复用的判据。

简而言之,数据评估的本质是数据可复用性评估,涉及数据质量评估、数据信任评估、数据文档质量评估研究的相关判据,但不仅限于这些相关主题的判据。对数据可复用性评估的判据综合后发现,其判据来源于数据获取平台(数据/数据存储库的可获得性、易获得性等)、数据本身(数据质量、数据的研究方法设计等)、数据产生的情境(元数据/数据描述信息、数据文档质量等)、数据生产者(伦理、声誉、能力及诚实度等)、数据复用者(专业知识、技能和经验等)及其他(如发表压力)。其中,数据质量评估是对数据可复用性评估的数据本身维度进行评估;数据信任评估涉及数据可复用性评估的数据本身(数据质量)、情境(数据文档质量)、数据生产者(声誉)等维度;数据文档质量评估是对数据可复用性中的情境维度进行评估。

2.3 科研人员数据复用的态度、意愿、行为及其影响因素

在纳入分析的 39 项文献中,有 25 项涉及科研人员数据复用感知数据、复用者的感知、意愿、行为及其影响因素的研究,其中与数据评估判据研究重复的有 11 篇,这部分重复的文献主要涉及数据复用行为中的机构/技术影响因素。我们将科学数据复用涉及的影响因素归并到 9 项分析主题中,并利用社会生态模型[65]作为影响因素的组织框架,将 9 项分析主题分别归为个人因素、机构/技术因素、学科/社会环境因素以及其他因素中,如表 4 所示。

2.3.1 个人因素

在很多学科领域(社会科学、STEM、工程学等),数据复用的态度或意愿不仅受到感知利益、感知关注(感知冒险)、感知努力、主观规范等的影响[29,31,35-36,57-58],还会受到年龄、资历、领域专长等的影响[23-24,38,66]。

(1)感知利益

"感知利益"是指数据复用者认为数据复用可以给科学研究带来价值或者节省时间、精力、花费等,在验证感知利益与数据复用意图或复用行为之间关系的定量研究中,得出了一致的结论,即感知利益与数据复用意图或行为呈现正相关[29,31,35-36,58-59]。可以得出结论,感知利益是数据复用之前首先要考虑的因素之一。

(2)感知关注

"感知关注"(感知冒险)指的是数据复用者认为数据复用会对数据知识产权、保密性带来风险或产生数据误用等,感知关注(感知冒险)

表 4 数据复用感知、意愿及行为影响因素

影响因素	相关概念(分析主题)	原文描述举例
个人因素	年龄、资历(专家、新手)	①younger researchers think more favorably about data sharing and reuse[66]
	信息需求(相关性)、数据处理技能、领域知识/领域专长	①users' incentives to use secondary data mostly depend on how well the data fit their information needs[58]:71 ②Specific knowledge aboutwho is working in what areas provide ecologists with insightsinto the types of data that are available for re-use[32]
	感知利益(感知有用性)、感知关注(感知冒险)、感知努力、主观规范、态度信仰	①perceived usefulness and internal resources were found to have positive relationships with scientists' data reuse intentions[31] ②Interviewee #3 said: "The data are there, they are already collected, it saves a lot of work, a lot of expense[59]:71 ③Perceived concern involved in data reuse negatively affects a health scientist's attitude toward data reuse[58] ④Attitude, norm of data reuse, and perceived effort significantly affect data reuse intention.[29]
机构/技术因素	数据可获得性、数据存储库/机构库的获得性,数据容易检索、易于访问、内部资源可用性、易用性,数据可靠性和重现性	①I would use other researchers' datasets if their datasets were easily accessible[66] ② At the discipline-level, theavailability of data repositories wasfound to have a significant positive relationship with scientists' data reuse intentions[31] ③resources at scientists' organizations positively influence scientists' intentions to reuse data[31]
	情境(Context)因素的可获得性(数据的预处理信息、来源信息、数据描述信息),数据文档的可获得性和完备性,元数据标准的可获得性和完备性	①Only by making existing data visible and accessible, and by providing the necessary contextual support, will researchers consider data re-use first[39] ②All case studies saw the need to address issues of documentation, context, and provenance as key requirements for the future re-use of data.[63]
学科/社会环境因素	数据类型、易用性,数据格式、软件或特殊的分析程序,学科本身的数据特征	①we found that reported use of models and remote-sensed data had a large positive effect on reuse behavior[36] ②I08 was also unable to open data because she was not sure of the format of the data or if she needed a special program to open and run the data.[26]
	学科规范、学科传统、学科风气、学科接受能力、学科领域文化、研究气候、学科接受能力、社会环境、社会规范	①disciplinary climate positively affects a health scientist's intention to reuse other scientists' data[58] ②interviewees disclosed two aspects related to their social environment they consider important when deciding whether to re-use data or not: their discipline and their peers[35]

续表

影响因素	相关概念(分析主题)	原文描述举例
其他因素	数据保密性、数据所有权、软硬件许可、高度专业化的技术使用;保护隐私、出版担忧	①The major barriers to re-use were found to be confidentiality, the importance of data ownership to status in the research community, hardware and software licences, and the use of highly specialized technology[39]
	组织支持:指导和培训(构建数据复用文化)、组织支持和协助的可获得性、数据生产者是否可联系到	① organizational support positively affects a health scientist's intention to reuse other scientists' data[26] ②I really needed a lot of external support for the data preparation and data analysis process[35]

对数据复用意图呈现负相关[29,31,35,58],但对数据复用的态度影响没有得到验证[57]。

(3)感知努力

"感知努力"指的是科研人员数据复用过程中,获取其他科研人员的数据并处理这些数据需要的时间和精力。大多数研究假设的预设是感知努力与数据复用意图呈现负相关,但这项结论在很多研究中没有得到验证。如社会科学[35]、工程学[59]、多学科领域[31],在纳入的文献中只有 Yoon、Kim 的研究验证了社会科学家的感知努力与数据复用意图呈现负相关[29],在社会科学研究中,二次数据分析需要大量的努力[22,26,61]。与社会科学中的数据相比,工程数据往往更系统化和结构化,更容易获取和使用。因此,工程科研人员对数据复用的感知努力可能不会在很大程度上影响工程研究人员的态度和意图。在 Kim、Yoon 对多学科的研究中,除了感知努力,其他因素如感知有用性、数据库的可获得性都与数据复用存在正相关关系[31],说明科研人员并没有在意数据复用过程中多付出努力,只要数据是可获得的而且是有用的,科研人员就愿意进行数据复用。

(4)主观规范

主观规范是指个人对于是否采取某项特定行为所感受到的社会压力。研究表明,科研人员支持数据复用的主观规范与数据复用态度或意图呈正相关关系[29,58-59],反对数据复用的主观规范与数据复用行为呈现负相关关系[36]。因多项研究表明,数据复用态度与数据复用意图或数据复用行为之间存在显著的正相关关系[29,35,58],可以认为这两种结论在本质上是一致的,即支持数据复用的主观规范与数据复用行为呈正相关。

(5)年龄、资历、领域专长

数据复用者的年龄和资历作为数据复用行为的影响因素得出了一致的结论,即年轻的科研人员虽然对数据复用的感知更为敏感,但比年长者对数据复用持有更谨慎和保守的态度,也比年长者提供更少的可复用数据[24,30,66]。领域专长对数据复用的影响因学科而存在差异,在考古学和生态学领域,数据复用受到领域知识和领域专长的影响,具有领域专长的研究人员可以帮助数据复用者更好地进行数据评估,但生物医学领域 3/4 的受访者认为领域专长不那么重要[38]。

2.3.2 机构/技术因素

机构/技术相关的因素指的是机构内是否有相应的基础设施获取数据,数据是否可获取,元数据标准是否可用等。

(1)可获得性(Accessability/Availability)

可获得性表示两方面涵义:一是数据是否可检索到、是否易于访问;二是是否有数据存储库可供使用。数据是否可以获取到是数据复用的前提条件,特别是数据存储库,它作为一种科学研究的网络基础架构,数据可获得性/数据存储库的可获得性/可访问性是数据复用过程中

面临的主要挑战之一[10,11,28,64,67],它可以培育学科的数据复用规范和文化[10,67],增强数据复用者的信任[28]。但其对数据复用意愿的影响还未得到一致的结论,如在工程学、多学科领域数据存储库的可获得性与数据复用意图呈现正相关[31,59],而在健康科学、社会科学领域,数据存储库的可获得性对数据复用意图的影响没有得到证实[29,58]。

(2)数据文档、元数据信息的可获得性

大量的研究文献表明,与情境因素相关的数据文档可获得性和完备性[24,26,33,37,63]、元数据/数据描述信息[23,37,63-64]]等都是数据复用的重要影响因素,正如 Faniel 等人对社会科学领域的研究表明,数据相关性、数据文档的质量等指标与数据复用的满意度有着显著的正相关关系[22];但在工程学领域,元数据标准的可获得性与数据复用意图之间的关系没有得到验证[59],这可能与工程学领域元数据标准刚刚开始使用有关,还需要继续验证。

2.3.3 学科/社会环境因素

(1)学科的数据类型和特征

不同学科的数据类型、数据格式、数据收集方法以及数据的易用性等属于数据本身的特征,也会影响数据复用行为[11,23,26,28,33,36],如 Curty 在研究中将数据的类型作为控制变量,划分为生物模型和遥感数据、自然科学调查与观察数据、社会科学访谈和调查数据三种类型,探究数据类型是否会影响数据复用行为,结果表明,只有生物学领域的模型和遥感数据与数据复用行为呈现正相关关系[36]。生物模型和遥感数据与其他领域收集的数据相比,是通过自动设备(如卫星)收集的,使用稳健的信息技术抓取和管理,这种数据比其他领域人工收集数据更容易复用。

(2)学科规范和学科文化

学科规范、学科传统、学科风气、学科接受能力、学科领域文化、研究气候、社会环境、社会规范等都是对学科环境的描述,指的是科研人员认为数据复用在他们的研究学科或社区中普遍存在的程度,以及在学科内是否有鼓励科学数据复用的文化等。学科规范或学科文化在不同的学科呈现出两种倾向:鼓励或限制。如果是鼓励数据复用则与数据复用意愿呈现正相关,反之则负相关。如在健康科学领域,科研人员正面临一种新的研究环境,即通过将多维临床数据与实验室数据相结合,增加数据分析能力,这种学科环境或文化将正向影响科研人员数据复用的意愿[58],同样在社会学领域,社会规范、学科气候与数据复用意图呈现正相关关系[29,35];而在教育学领域,政府官员在教育政策的制定过程中,并不重视科研人员数据分析的结果,导致科研人员对数据复用产生怀疑[68]。

2.3.4 其他因素

除了以上三方面的影响因素,数据本身、组织支持与帮助、特定数据类型的出版压力等也会影响数据复用。如保密性、数据所有权等是工程科研人员数据复用的主要障碍因素[39];组织的指导和培训[59]、组织支持和帮助[26]、数据生产者是否可联系到[26,35]等也是影响数据复用的影响因素,并且已有研究表明,组织支持或内部资源的可获得性对数据复用意愿产生正向影响[31,58];对于定性数据的复用,成果出版压力会影响数据复用的意愿,因为定性数据研究人员表示很难出版二次数据分析的研究成果[28]。

3 讨论与研究展望

3.1 研究方法

科研人员数据复用行为研究延续了信息行为研究中的两个重要方法:访谈和调查。基于访谈的质性研究可以探索诠释研究对象的潜在特征,基于调查的量化研究可以大规模地解释普遍现象,这两种研究方法也是研究用户信息行为最常用的方法[69]。这两种方法在科研人员数据复用态度、意愿和行为的研究中依赖于科研人员的感知和自我报告,而自我报告的复用

态度、意愿和行为不能代替实际数据复用行为，因此要研究数据复用行为，应该进一步拓宽研究方法，比如实验方法中的准实验出声思维法，能够有效地再现思维过程，克服访谈法、问卷调查法的内省性和反馈性缺点，可以更精确地捕捉到研究对象在特定环境中的感知变化。

除了利用实验法，学者们还基于文献的计量、内容分析等非介入性（Unobtrusive）方法揭示科学数据复用行为。如 Zhang 在其博士论文中使用内容分析法分析文献中的数据使用模式[70]；Cheng 利用定性内容分析法，对医学开放数据中的 NHIRD 数据集的使用阶段特征和使用行为进行分析[71]；Piwowar 等利用文献计量法，研究得出数据可公开获得性独立于期刊影响因子、出版日期和作者原国籍，与引文增加显著相关[72]；Park、Wolfram 研究了遗传基因的 148 篇引用文章，以确定影响数据共享和复用的因素[73]。因此，基于文献的计量分析和内容分析法可以更客观地探究不同学科数据复用行为及其影响因素，是对科研人员数据复用行为研究方法的有效补充。

3.2 研究内容

（1）现有研究以理性行为理论和计划行为理论为主构建科研人员数据复用影响因素作用机制模型，模型中涉及的构念还不够细致和深入，个人因素仅讨论了个人感知和人口计量的一些构念，对科研人员的心理和社会特征的挖掘不够。上述讨论中得出人际关系是重要的可复用数据源，受此启发，可在影响因素作用机制模型中加入“关系”维度；技术相关因素仅仅讨论了数据存储库的可获得性构念，情境相关的构念（比如数据文档的可获得性、元数据的完备性等）在模型中体现很少。另外，数据复用的意愿和行为涉及多个层面的因素，主要有个人因素、技术因素、社会/学科因素等，这些因素如何共同影响数据复用，需要进一步采用整体和多维的方法进行验证。

（2）现有研究集中在数据获取来源、数据评估判据、数据复用感知、意愿和行为影响因素方面，对于数据复用过程中的数据需求、数据检索与获取、数据处理与使用等方面的研究则比较欠缺；学科之间的数据复用过程差异很大，应根据不同学科数据需求、数据检索与获取、数据处理和数据评估特点，进一步挖掘不同学科的数据复用流程并建模。

（3）数据的可复用性评估作为科研人员数据复用流程中的核心阶段，尽管有多位学者直接或间接地探讨了数据的可复用性，但其内涵还没有一致的结论。从整体视角理解基于数据复用者感知的数据可复用性则是一个非常复杂的问题，涉及数据可获得性（如数据可获取性、数据存储库可获得性）、数据本身（如数据质量）、数据情境信息（如相关性）、数据生产者（声誉、伦理等）等维度。因此，如果对科研人员感知的数据可复用性建模的话，就应该包含以上维度中的核心要素，并厘清数据可复用性与数据质量、数据信任、数据文档质量等概念的关系，比如数据可获得性构念在数据质量判据中位居第二位，是较为重要的判据之一，而在数据复用的影响因素中数据可获得性与数据复用意图呈现正相关关系，数据可获得性究竟是直接影响数据复用意愿还是通过数据质量间接影响数据复用的意愿，需要进一步验证。另外，也有学者指出，要理解数据可复用性的内涵，主要可以从政策、法律、经济和技术四个层面来考虑。技术层面，其理论基础来源于关系思维、知识边界、数据表征等；非技术层面，涉及数据作者和数据用户之间的关系等[74]。

（4）数据质量评估是数据复用过程中的重要环节，数据质量的研究起源于信息质量[75-76]和管理信息系统的研究[77-78]，前期的信息质量研究为数据质量研究奠定了较好的基础，数据质量本身具有高度的情境敏感性和领域独立性，应根据学科领域，理解不同利益相关者的数据质量感知和需求，开发面向特定用户群体的数据质量模型[5,20,57]。

（5）科学数据复用的情境是科研人员数据

相关性评估的重要判据。元数据作为传达情境信息的理想工具,其完备性对科研人员数据评估有重要影响,但目前结构化元数据仅仅提供数据的技术、结构以及方法要素[23],是对实验过程的不完全抽象,还不能提供足够的情境信息[79],要完全理解数据的含义,还应包含数据质量指标、实验设计文件[12]、数据产生的独特细节[6,23]、数据提供者和数据存储库的声誉等。然而数据生产者根本不可能记录他们对数据的每一个决定,以及他们在数据创造中使用的隐性知识[33,61,80],而且很难知道哪些情境信息能够满足用户需求和期望[63],因此,需要对不同学科数据评估的情境框架进一步探究,辅助元数据设计。

(6)科学数据复用过程中,新兴的科学数据共享平台或者机构知识库的可获得性是影响数据复用的重要因素,然而对科学数据平台的可用性研究还比较薄弱。目前仅仅从平台的数据质量、元数据的完备性等探讨其可用性,对于从人机交互视角(比如数据的访问速度、导航的易用性、在线支持、响应时间等)探讨可用性的研究还需要加强。

(7)数据共享是数据复用的前提,数据复用是数据共享的目的。从数据生命周期来看,数据共享与数据复用分别位于数据生命周期的不同阶段,是一个有机整体。从数据共享与复用过程来看,数据共享涉及数据处理、保存等,是后续数据复用过程中数据检索获取、评估使用的前提;从数据共享与复用的意愿、行为及影响因素上看是存在相互关系的,比如从技术的角度看,数据存储库是数据的可共享性与数据的可复用性的共同影响因素,并且有研究表明数据复用经验显著影响科研人员对数据共享的感知和共享规范[81]。因此,应将数据共享与复用放在同一研究视野中,研究科研人员数据共享与复用行为。

目前,科研人员数据复用研究在国外已经出现了7篇博士论文,而国内鲜有系统化的研究,非常有必要就此议题进行本土化探索,对不同学科的数据复用流程、数据可复用性评估、数据质量评估等进行建模,并从不同的理论视角挖掘科学数据复用行为及其影响因素作用机制,运用多种研究方法进行多层次、细粒度的研究,更全面地揭示我国科研人员的数据复用行为。

4 结论

本文利用系统综述和文献元综合方法对科研人员数据复用的实证文献进行了细粒度的揭示,客观评述已有的研究贡献,并对主要研究主题的研究结论进行批判性解释综合。结果显示,科研人员数据复用研究涉及自然科学、社会科学、生物医学等几乎所有学科,科学数据的获取来源以纸质文献、学科知识库和数据文档为主;科学数据评估包括数据可复用性评估、数据信任评估、数据质量评估、文档质量评估等,这些概念具有高度的情境敏感性和领域独立性,不同学科科研人员在数据复用的过程中关注的评估要素存在差异,但几乎所有学科的科学数据复用过程都是复杂的,是一个类似知识复用[6]、非线性的[1]、耗时费力的迭代过程[80]。科研人员数据复用的影响因素主要涉及个人因素、机构/技术因素、学科/环境因素以及其他相关因素,这些影响因素也会随着学科和领域的不同而存在差异。因此,采用"一刀切"的解决方案来理解数据复用行为是不现实的,应根据学科领域,对不同学科的数据复用流程进行建模,厘清数据可复用性的判据之间的关系,并从不同的理论视角采用整体和多维的方法对数据复用的影响因素进行实证研究。

本文仍存在一些缺陷:①本文采用的文献元综合方法在国内图情领域是首次使用,在研究问题综合的过程中使用批判解释综合的方法,使用这种方法的前提是在同一研究主题内,有大量相同概念或相同构念的研究,由于数据复用研究是一个较新的研究主题,特别是数据复用意愿或态度的研究仅仅有6篇定量实证研究,这就不得不辅以一些定性研究结论来解释

影响因素,但在实际的综合过程中,发现将定量和定性研究的结论进行综合确实有些难度,因此对元综合方法的首次使用是否存在理解偏差还需要未来学者的考证。②尽管搜索策略尽可能地穷尽文献中涉及的检索词,但结构化搜索策略仍不能避免检索命中率偏少,我们通过浏览主要期刊、参考文献滚雪球、引文追踪等方法来弥补这些缺陷,但仍有部分文献遗漏;另外,检索词以"Data"为核心词进行检索,未考虑关于数据的定义[82]中多种数据形式的检索词,比如文本、数字、图像、视频或电影、音频、软件、算法、方程式、动画、模型等在不同学科都可能是数据的表现形式,比如生物医学中的神经科学和放射学,其数据主要是影像(Imaging),如果仅仅以"Data"为检索词不可避免对一些数据复用研究有遗漏。

参考文献

[1] Yoon A. Data reuse and users' trust judgments: toward trusted data curation[D]. Chapel Hill, NC: University of North Carolina at Chapel Hill Graduate School,2015.

[2] European Commission. Guidelines on open access to scientific publications and research data in Horizon 2020 [DB/OL].[2018-5-20].https://www.openaire.eu/guidelines-on-open-access-to-scientific-publications-and-research-data-in-horizon-2020.

[3] The National Institutes of Health. Data sharing policy and implementation guidance[EB/OL].[2018-05-20].https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm.

[4] National Science Board. Digital research data sharing and management[EB/OL].[2018-05-20].https://www.nsf.gov/nsb/publications/2011/nsb1124.pdf.

[5] Huang H. Perception of quality in genome annotation work[D]. Tallahassee, Florida: The Florida State University,2010.

[6] Murillo A P. Data sharing and data reuse: an investigation of descriptive information facilitators and inhibitors [D]. Chapel Hill, North Carolina: University of North Carolina at Chapel Hill,2016.

[7] Sands A E, Borgman C L, Wynholds L, et al. Follow the data: how astronomers use and reuse data[C/OL]. [2018-05-01].https://www.asis.org/asist2012/proceedings/Submissions/341.pdf.

[8] Faniel I M, Barrera-Gomez J, Kriesberg A, et al.A comparative study of data reuse among quantitative social scientists and archaeologists[C]//iConference 2013 Proceedings, Philadelphia,PA: iSchools,2013:797-800.

[9] Bishop L, Kuula-Luumi A. Revisiting qualitative data reuse[J/OL].SAGE Open, 2017,7 (1)[2018-04-24]. http://journals.sagepub.com/doi/full/10.1177/2158244016685136.

[10] Kriesberg A, Frank A, Faniel I M, et al.The role of data reuse in the apprenticeship process[C/OL]//Grove A. Proceedings of the 76th asis&t annual meeting: beyond the cloud: rethinking information boundaries. Montreal, Canada: ASIST, Wiley,2013:50[2018-04-25]. http://www.asis.org/asist2013/proceedings/openpage.html.

[11] Wallis J C, Rolando E, Borgman C L. If we share data, will anyone use them? data sharing and reuse in the long tail of science and technology[J]. PLoS ONE,2013,8(7):e67332.

[12] Birnholtz J, Bietz M. Data at work: supporting sharing in science and engineering[C]//Pendergast M, Schmidt K, Simone C, et al. Proceedings of the 2003 international ACM SIGGROUP conference on supporting group work. Sanibel Island, FL: GROUP'03,2003:339-348.

[13] Borgman C L. The conundrum of sharing research data[J]. Journal of the American Society for Information Science and Technology,2012(63):1059-1078.

[14] National Academy of Science. Ensuring the integrity, accessibility, and stewardship of research data in the digital age[EB/OL].[2018-05-20].http://www.nap.edu/catalog.php?record_id=12615.

[15] Whyte A, Pryor G. Open science in practice: researcher perspectives and participation[J].International Journal of Digital Curation,2011(6):199-213.

[16] Corrall S, Kennan M A, Afzal W. Bibliometrics and research data management services: emerging trends in library support for research[J].Library Trends,2013,61,(3):636-674.

[17] Tenopir C, Sandusky R J, Allard S, et al. Research data management services in academic research libraries and perceptions of librarians[J].Library & Information Science Research,2014, 36(2):84-90.

[18] Coates, H. Ensuring research integrity: the role of data management in current crises[J].College & Research Libraries News, 2014,75(11):598-601.

[19] CUL Data Working Group, Cornell University Library. Digital research data curation: overview of issues, current activities, and opportunities for the cornell university library[EB/OL].[2018-05-24].http://ecommons.cornell.edu/bitstream/1813/10903/1/DaWG_WP_final.pdf.

[20] Stvilia B, Hinnant C C, Wu Shuheng, et al. Research project tasks, data, and perceptions of data quality in a condensed matter physics community[J]. Journal of the Association for Information Science and Technology, 2015,66(2):246-263.

[21] Faniel I M, Kansa E, Whitcher K S, et al. The challenges of digging data: a study of context in archaeological data reuse[C]//Proceedings of the 13th ACM/IEEE-CS joint conference on digital libraries, NY, USA: ACM, 2013: 295-304.

[22] Faniel I M, Kriesberg A, Yakel E. Social scientists' satisfaction with data reuse[J].Journal of the Association for Information Science & Technology,2016, 67(6):1404-1416.

[23] Faniel I M, Jacobsen T E. Reusing scientific data: how earthquake engineering researchers assess the reusability of colleagues' data [J]. Computer Supported Cooperative Work, 2010, 19(3-4):355-375.

[24] Faniel I M,Kriesberg A, Yakel E. Data Reuse and sense making among novice social scientists[J]. Proceedings of the American Society for Information Science and Technology, 2012, 49(1):1-10.

[25] Yoon A.Visible evidence of invisible quality dimensions and the role of data management[C]//iConference 2016 Proceedings. Philadelphia,PA: iSchools ,2016.

[26] Yoon A.Red flags in data: learning from failed data reuse experiences[C]//Proceedings of the Association for Information Science and Technology, 2016,53(1):1-6.

[27] Yoon A. Data reusers' trust development[J].Journal of the Association for Information Science & Technology, 2017,68(4):946-956.

[28] Yoon A. "Making a square fit into a circle": researchers' experiences reusing qualitative data[C]// Proceedings of the77th ASIS&T Annual Meeting: Connecting Collections, Cultures, and Communities. Seattle, Washington: ASIST, Wiley,2014:51.

[29] Yoon A, Kim Y. Social scientists' data reuse behaviors: exploring the roles of attitudinal beliefs, attitudes, norms, and data repositories[J]. Library & Information Science Research,2017,39(3):224-233.

[30] Kim J, Schuler E R, Pechenina A. Predictors of data sharing and reuse behavior in academic communities[C]// Alemneh D G, Allen J, Hawamdeh S. Knowledge discovery and data design innovation: proceedings of the International Conference on Knowledge Management (ICKM 2017), Singapore: World Scientific, 2017:1-25.

[31] Kim Y, Yoon A. Scientists' data reuse behaviors: a multilevel analysis[J]. Journal of the Association for Information Science & Technology, 2017,68(12):2709-2719.

[32] Zimmerman A S. Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse[J]. International Journal on Digital Libraries,2007,7 (1-2): 5-16.

[33] Zimmerman A S. New knowledge from old data the role of standards in the sharing and reuse of ecologicaldata[J]. Science, Technology & Human Values, 2008,33(5):631-652.

[34] Zimmerman A S. Data sharing and secondary use of scientific data: experience of ecologists[D]. Michigan: The University of Michigan, 2003.

[35] Curty R G. Beyond "data thrifting": an investigation of factors influencing research data reuse in the social sciences[D]. Syracuse,NY: Syracuse University,2015.

[36] Curty R G, Crowston K, Specht A, et al. Attitudes and norms affecting scientists' data reuse[J]. PLoS ONE, 2017,12(12):e0189288.

[37] Daniels M G. Data reuse in museum contexts: experiences of archaeologists and botanists[D], Michigan :University of Michigan,2014.

[38] Federer L M, Lu Y-L, Joubert D J, et al. Biomedical data sharing and reuse: attitudes and practices of clinical and scientific research staff[J]. PLoS ONE,2015,10(6): e0129506.

[39] Howard T, Darlington M, Ball A, et al. Opportunities for and barriers to engineering research data re-use[R]. Bath, UK: University of Bath, 2010.

[40] Catalano A. Patterns of graduate students' information seeking behavior: a meta-synthesis of the literature[J].Journal of Documentation, 2013,69(2):243-274.

[41] Ankem K. Evaluation of method in systematic reviews and meta-analyses published in LIS[J].Library and Information Research, 2008,32(101):91-104.

[42] Brettle A. Systematic reviews and evidence based library and information practice[J].Evidence based Library and Information Practice, 2009,4(1): 43-50.

[43] Saxton M L. Meta-analysis in library and information science: method, history, and recommendations for reporting research[J]. Library Trends, 2006,55(1):158-170

[44] Urquhart C. Meta-synthesis of research on information seeking behaviour[J].Information Research,2011,16(1).

[45] Urquhart C, Yeoman A. Information behaviour of women: theoretical perspectives on gender[J]. Journal of Documentation, 2010,66(1):113-139.

[46] Urquhart C. Systematic reviewing, meta-analysis and meta-synthesis for evidence-based library and information science[J]. Information Research, 2010,15(3) .

[47] Sackett D L, Haynes R B.On the need for evidence-based medicine[J]. Evidence-Based Medicine,1995,1:5-6.

[48] Crowe M, Sheppard L. A review of critical appraisal tools show they lack rigor: alternative tool structure is proposed[J]. Journal of Clinical Epidemiology,2011,64:79-89.

[49] Koufogiannakis D, Booth A, Brettle A. ReLIANT: reader's guide to the literature on interventions addressing the

need for education and training[J]. Library and Information Research,2006,30(94):44-51.

[50] Glynn L. A critical appraisal tool for library and information research[J]. Library Hi Tech, 2006,24(3): 387-399.

[51] EBL Critical Appraisal Checklist[EB/OL].[2019-03-15].http://ebltoolkit.pbworks.com/f/EBLCriticalAppraisalChecklist.pdf.

[52] Israel G D. Determining sample size[EB/OL].[2018-05-01].https://www.tarleton.edu/academicassessment/documents/Samplesize.pdf.

[53] Hamilton M B. Online survey response rates and times: background and guidance for industry[EB/OL].[2018-04-12]http://www.supersurvey.com/papers/supersurvey_white_paper_response_rates.pdf.

[54] Dixon-Woods M, Cavers D,Agarwal S, et al. Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups[J].BMC Medical Research Methodology,2006(6):35.

[55] Flemming K.Synthesis of quantitative and qualitative research: an example using critical interpretive synthesis[J]. Journal of Advanced Nursing,2010,66(1):201-217.

[56] Heaton J, Corden A, Parker G. Continuity of care': a critical interpretive synthesis of how the concept was elaborated by a national research programme[J].International Journal of Integrated Care,2012,12(2):1-9.

[57] Huang H, Stvilia B, Jorgensen C, et al. Prioritization of data quality dimensions and skills requirements in genome annotation work[J]. Journal of American Society of Information Science and Technology, 2012,63(1):195-207.

[58] Joo S H, Kim S J, Kim Y. An exploratory study of health scientists' data reuse behaviors: examining attitudinal, social, and resource factors[J].Aslib Journal of Information Management, 2017,69(4):389-407.

[59] Joo Y K, Kim Y. Engineering researchers' data reuse behaviours: a structural equation modelling approach[J]. The Electronic Library,2017,35(6):1141-1161.

[60] Schmidt B, Gemeinholzer B, Treloar A. Open data in global environmental research: the belmont forum's open data survey[J]. PLoS ONE, 2016,11(1):e0146695.

[61] Niu J F.Perceived documentation quality of social science data[D]. Michigan :The University of Michigan, 2009.

[62] Wynholds L, et al. When use cases are not useful: data practices, astronomy, and digital libraries[C]. Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries,2011:383-386.

[63] Carlson S, Anderson B. What are data? the many kinds of data and their implications for data re-use[J]. Journal of Computer-Mediated Communication, 2007,12(2):635-651.

[64] Shen Y. Data Sustainability and reuse pathways of natural resources and environmental scientists[J/OL]. New Review of Academic Librarianship,2018,24(2):136-156 [2018-04-13].https://export.arxiv.org/ftp/arxiv/papers/1803/1803.01788.pdf.

[65] Stokols D. Translating social ecological theory into guidelines for community health promotion[J]. American Journal of Health Promotion,1996,10(4):282-298.

[66] Tenopir C, Dalton E D, Allard S, et al. Changes in data sharing and data reuse practices and perceptions among scientists worldwide[J]. PLoS ONE, 2015,10(8): e0134826.

[67] Shen Y. Research data sharing and reuse practices of academic faculty researchers: a study of the virginia tech data landscape[J]. International Journal of Digital Curation, 2015, 10(2):157-175.

[68] 林奇秀,赖璟毅.台湾社会科学学者资料再用行为之研究[J].图书资讯学研究, 2017,11(2):95-138.(Lin

C S, Lai C Y. Data reuse behavior among Taiwan social scientists[J]. Journal of Library Information Science Research,2017,11(2):95-138.)

[69] Julien H, O'Brien M. Information behaviour research: where have we been, where are we going?[J]. Canadian Journal of Information and Library Science, 2014,38(4):239-250.

[70] Zhang J. 2011data use and access behavior in escience—exploring data practices in the new data-intensive science paradigm[D]. Philadelphia: Drexel University,2011.

[71] Cheng W C, Chiu M H P. How do medical researchers use open health data? a case study on data reuse behavior of using NHIRD in Taiwan[C].Proceedings of the Association for Information Science and Technology, 2017, 54(1):637-639.

[72] Piwowar H A, Day R S, Fridsma D B.Sharing detailed research data is associated with increased citation rate[J]. PloS ONE,2007,2(3): e308.

[73] Park H, Wolfram D. An examination of research data sharing and re-use: implications for data citation practice [J]. Scientometrics,2017,111(1):443-461.

[74] Thanos C. Research data reusability: conceptual foundations, barriers and enabling technologies[J/OL]. Publications,2017,5(1):1-19[2018-04-23]. http://www.mdpi.com/2304-6775/5/1/2/htm.

[75] Arazy O, Kopak R. On the measurability of information quality[J]. Journal of the Association for Information Science and Technology, 2011,62(1): 89-99.

[76] Knight S, Burn J. Developing a framework for assessing information quality on the world wide web[J]. Informing Science Journal,2015(8):159-172.

[77] Madnick S, Lee Y. Editorial for the inaugural issue of the *ACM Journal of Data and Information Quality*[J].ACM Journal of Data and Information Quality, 2009,1(1):1-6.

[78] 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.

[79] Kern D, Mathiak B. Are there any differences in data set retrieval compared to well-known literature retrieval? [C]//Kapidakis S, Mazurek C, Werla M. Research and advanced technology for digital libraries. lecture notes in Computer Science. Springer, Cham.2015(9316).

[80] Rolland B, Lee C P. Beyond trust and reliability: reusing data in collaborative cancer epidemiology research [C]//Bruckman A, Counts S, Lampe C, et al. Proceedings of the 2013 Conference on Computer Supported Cooperative Work, New York, NY, USA: ACM ,2013: 435-444.

[81] Kim Y, Nah S. Internet researchers' data sharing behaviors: an integration of data reuse experience, attitudinal beliefs, social norms, and resource factors[J]. Online Information Review ,2018,42(1): 124-142.

[82] National Science Board.Long-lived digital data collections enabling research and education in the 21st century[R/OL].[2018-05-01].https://www.nsf.gov/nsb/meetings/2005/LLDDC_draftreport.pdf.

孙玉伟 南京大学信息管理学院博士研究生,山东师范大学副研究馆员。江苏 南京 210023。

成 颖 南京大学信息管理学院教授,博士生导师。江苏 南京 210023。

谢 娟 南京大学信息管理学院博士研究生。江苏 南京 210023。

(收稿日期:2018-05-23;修回日期:2019-03-28)