

教育学领域学者数据重用行为及特征研究

——基于基金项目论文的内容分析

姜鑫妍 王翠萍 宋雯琪

(东北师范大学信息科学与技术学院, 吉林 长春 130117)

摘要: [目的/意义] 为了解我国教育学领域学者数据重用行为特征, 帮助图书馆等机构在其科学数据管理和服务过程中制定更有针对性的数据服务策略。[方法/过程] 本研究运用内容分析法对我国 2017—2018 年 544 篇基金论文中的科学数据重用行为进行分析。[结果/结论] 研究发现, 教育学领域学者数据重用比例较高且较为稳定, 政府网站、数据中心及期刊论文是教育学领域学者重用数据的主要来源渠道, 但存在着科学数据二次分析较少及数据重用不规范的问题。

关键词: 科学数据; 数据重用; 数据重用行为; 教育学; 内容分析法

DOI:10.3969/j.issn.1008-0821.2024.04.008

[中图分类号] G316 [文献标识码] A [文章编号] 1008-0821 (2024) 04-0077-10

Research on Data Reuse Behavior and Characteristics of Scholars in Education

——Content Analysis Based on Fund-sponsored Papers

Jiang Xinyan Wang Cuiping Song Wenqi

(College of Information Science and Technology, Northeast Normal University, Changchun 130117, China)

Abstract: [Purpose/Significance] The paper aims to discovery the characteristics of data reuse behavior of scholars in education, and helps libraries and other institutions develop more targeted data service strategies in the process of scientific data management and service. [Method/Process] The paper used content analysis to analyze the scientific data reuse behavior in 544 fund-sponsored papers from 2017 to 2018. [Results/Conclusions] The paper finds that data reuse ratio of scholars in education is relatively high and stable, and it is more inclined to reuse survey data. In terms of data reuse sources, government websites, data centers and journal papers are the main sources of data reuse for scholars in education, and there are some problems such as less secondary analysis of scientific data and data reuse specifications are inadequate.

Key words: scientific data; data reuse; data reuse behavior; education; content analysis

科学数据是科学研究的基础, 在开放科学的背景下, 国内外陆续发布了众多促进科学数据开放利用的相关政策和规范指南。近年来, 我国科学数据规模增长迅速, 但存在数据利用率不高的问题。随着数据密集型研究范式的到来, 数据重用研究热度逐渐升高。数据重用(Data Reuse), 也称为数据复用、数据再利用, 是对科学数据的重复利用及二次

分析。本文所指的数据重用定义采取一种较为宽泛的概念, 即数据重用是数据的重新使用或重复利用, 无论之前是否使用过该数据, 只要研究中所使用的数据或数据集不是科研人员自行收集的原始数据, 其数据使用或数据分析过程就应视为数据重用行为。科学数据重用行为的主要表现为数据重用目的、原始数据特征、数据处理分析、数据重用规范等方面。

收稿日期: 2023-07-05

作者简介: 姜鑫妍 (1999-), 女, 硕士研究生, 研究方向: 信息资源管理。王翠萍 (1974-), 女, 教授, 研究方向: 信息组织与服务、信息分析与咨询。宋雯琪 (1998-), 女, 硕士研究生, 研究方向: 信息资源开发与管理。

在科技文献中可表现为数据使用、数据引用、数据来源提及等。

国外较早开始研究数据重用相关问题,研究主题较为广泛和分散,大致可总结为数据重用概念及重要性的研究^[1-7],元数据与数据重用的研究^[8-9],数据重用行为及态度的研究^[10-17]以及其他相关研究^[18-19]。国内对于数据重用的研究主要集中对数据重用行为特征分析^[20-27]、数据重用的影响因素研究^[27-36]、构建数据重用模型的相关研究^[37-41]以及探讨数据权益问题^[42-43]。目前对于学界内针对数据重用的研究更多地关注较为具体的问题,有部分研究从不同学科论文中的数据引用实践入手,分析了生物学^[44]、化学^[45]、生命科学、地球物理学^[46]、人口学^[47]等学科领域的论文引用特征。在现有研究中,大多数学者主要聚焦于各类型数据较多的自然科学等学科领域的论文数据重用现象的研究,而对人文社科领域内学者们对数据重用现象研究较少。教育学作为社会科学领域中的一个重要范畴,在其研究领域内学者们广泛应用观察法、调查法等实证方法产生了大量的科学数据,同时普及了数据重用的意识和实践。经基金资助的论文成果要求进行科学数据汇交,科学数据管理相对更规范,并且基金论文具有阶段性特点能很好地反映科研人员数据重用行为的阶段性特征,与本文研究目的更为契合。因此,本文选取教育学领域基金论文,运用内容分析法试图揭示教育学领域的论文数据重用行为特征,以及数据重用各阶段特点,以期在一定程度上为图书馆、数据出版商、数据中心等机构开展数据服务提供参考。

1 样本选择与数据处理

1.1 样本选择

全国教育科学规划立项课题是教育学领域专门的国家级基金项目,能够反映我国教育科学研究的普遍态势,具有较强的代表性。由于科研基金立项项目具有长周期和持续性等特点,为充分了解学者们在其基金项目全周期内各阶段论文中的数据重用特征,本研究选择立项时间距今 5 年左右的基金项目产出的科研论文作为数据分析样本,最终确定了 2017 年和 2018 年全国教育科学规划项目数据为总体数据。

1.2 数据获取

本文从基金项目出发,统计了 2017 年及 2018 年立项课题数量,分别是 488 项和 551 项。由于全国教育科学规划项目各年基金立项总数相差不大,但各类别课题立项数量不一致,其中国家一般课题所占比重最大,国家重大及重点课题所占比例最低。不同类别基金项目论文数量差异明显,同一类别课题已发文数量相对一致。一般而言,一项国家一般项目发表论文数量多集中在 3~8 篇,国家重大及重点基金论文数量多为 10 篇左右。以往研究运用内容分析法分析论文数量并不一致,多则 1 000 余篇,少则一两百篇,但总体数量分布在 400~1 500 之间,因此通过抽样方法即可获得满意的数据量。为保证样本数据的代表性,抽样选择基金项目数量应不少于 80。考虑到各类基金项目数量分布的不均衡性,本文采取分层随机抽样方法,分年份基金类别随机抽取各类基金项目数量的 10%,总计抽取 100 项,其中 2017 年与 2018 年立项各类基金项目分别为 49 项和 51 项。在项目编号的规定方面,利用在线随机数生成器得到随机数,并依据此随机数对所抽样项目进行了标注。具体标注规则为:如被抽取课题已发表论文数量为 0 篇,则被抽取基金项目编号依次顺延,但是应保证顺延后序号与随机数不重合,2017—2018 年实际数据抽样结果如表 1 和表 2 所示。

根据上述抽样规则,本研究共抽样选择了 100 项课题,在中国知网检索上述被抽取的基金课题已发表的相关论文成果,共获取到研究论文 544 篇,平均每项课题发表篇数为 5.4 篇,数据量满足分析条件。

1.3 数据处理

1.3.1 数据分析方法

本文采用内容分析法从数据是否重用、数据分析与处理方式、数据重用规范等方面对我国基金项目论文中的科学数据重用行为进行分析,通过人工识别的方法对所抽取的课题论文中所提及的非科研人员自行收集的原始数据或数据集进行了提取。在参考已有研究成果并结合教育学学科数据特征后,经过反复修改,最终形成数据重用行为内容分析编码表,如表 3 所示。

表 1 2017 年基金项目抽样表
Tab. 1 2017 Fund Project Sample Table

2017 基金类别	抽样数(10%)	抽 样 编 号
国家重大(序号 1~4)	0	
国家重点(序号 5~12)	1	10
国家一般(序号 13~240)	23	15; 20; 23; 33; 39; 52; 75; 78; 81; 87; 103; 108; 119; 136; 157; 165; 177; 180; 188; 214; 216; 219; 233
国家青年(序号 241~279)	4	242; 248; 261; 275
教育部重点(序号 280~428)	15	297; 301; 306; 324; 326; 348; 351; 358; 363; 365; 367; 375; 405; 408; 425
教育部青年(序号 429~488)	6	430; 445; 454; 458; 473; 481
总计 488	49	

表 2 2018 年基金项目抽样表
Tab. 2 2018 Fund Project Sample Table

2018 基金类别	抽样数(10%)	抽 样 编 号
国家重大(序号 1~5)	1	1
国家重点(序号 6~16)	1	12
国家一般(序号 17~236)	22	21; 33; 47; 54; 65; 96; 108; 112; 122; 127; 136; 145; 166; 175; 179; 181; 194; 202; 209; 217; 232; 236
国家青年(序号 237~279)	4	248; 258; 259; 263
西部项目(序号 280~296)	2	281; 286
教育部重点(序号 297~447)	15	305; 313; 316; 322; 325; 352; 358; 371; 377; 379; 405; 414; 424; 433; 438
教育部青年(序号 448~511)	6	448; 477; 482; 492; 505; 507
总计 511	51	

1.3.2 数据编码说明

1) 基本要素编码

基金类别。全国教育科学规划课题是国家社科基金教育学单列学科课题，课题类别包括国家重大、教育部专项等 8 类。内容编码考虑基金类别主要是因为不同类别的课题发表的论文数量和研究方法存在差异，因此样本抽样区分基金类别可以使论文类型和数量分布更为均衡。

论文类型。本研究将已发表的基金论文分为实证论文和非实证论文两类。实证论文指的是通过观察、调查、实验等实证方法收集数据而写成的论文。

2) 数据重用

是否数据重用。如果“是”则标为“1”，“否”

记为“0”。

数据重用目的。本文所指的数据重用目的主要包括新的研究、佐证观点、研究结果比较以及用做背景资料(研究背景)4 大类。

3) 原始数据特征

原始数据类型。科学数据类型可按照研究方法分为调查数据和非调查数据。本文提到的调查数据包括统计数据在内，非调查数据是指调查数据以外的数据，包括观察数据、实验数据等多种类型。

数据重用来源渠道。数据来源渠道可分为正式和非正式两种方式。正式的数据来源渠道包括图书、期刊论文、年鉴等正式出版物及政府网站等其他官方统计网站或公开的调查报告等。非正式来源渠道有邮件、私人联系等。一般可以通过文中提及、参

表 3 数据重用行为内容分析编码表
Tab. 3 Data Reuse Behavior Content Analysis Coding Table

基本大类	一级类目	二级类目
基本要素 编码	基金类别	国家重大
		国家重点课题
		国家一般
		国家青年
		西部项目
		教育部重点
		教育部青年
		教育部专项
		实证
		非实证
数据重用 行为编码	论文类型	是否数据重用
		是(取值为 1)
		否(取值为 0)
		新的研究
		佐证观点
		研究背景
		研究结果比较
		数据类型: 调查数据
		非调查数据
		数据来源渠道: 期刊论文
数据重用 行为编码	原始数据特征	政府网站
		其他
		直接使用
		整理成图表
		其他
		数据重用格式: 参考文献
		文中提及
		其他
		数据元素: 必备元素
		可选元素
数据重用 行为编码	数据重用规范	数据重用位置: 引言
		正文
		结语
		数据重用障碍(如未知数据源)等
		其他
		数据重用障碍(如未知数据源)等
		数据重用障碍(如未知数据源)等
		数据重用障碍(如未知数据源)等
		数据重用障碍(如未知数据源)等
		数据重用障碍(如未知数据源)等

考文献、注释等方式得知数据来源信息。根据初步分析, 本文将数据重用来源渠道大致分为期刊论文、政府网站及其他三大类。

4) 数据处理与分析方式

本研究将数据处理与分析分为数据使用、数据统计分析、整理成图表和其他四大类(其他处理方式如文中明确提到借鉴、改编量表等)。

5) 数据重用规范

数据重用格式。本文科学数据重用规范分析的主要内容包括数据重用格式、数据元素、数据重用位置 3 个方面, 其中科学数据重用格式主要有参考文献、注释、文中提及以及其他 4 种形式, 除此之外还包括序号、标点符号的使用等较为细节的部分。

数据描述元素。可分为必备元素与可选元素。必备元素有创建者、标题、URL 等, 可选元素如 DOI、数据分级等。

数据重用位置。本文将数据重用位置分为引言、正文、结语 3 种, 虽然这种区分较为简单, 但可以与数据重用目的相对应。

2 教育学领域基金论文中的数据重用行为统计分析

2.1 数据重用目的

一般而言, 数据重用目的有 4 种: 新的研究、研究佐证、研究结果比较以及用做背景资料, 如图 1 所示。图 1 由图(a)、图(b)以及图(c)组成, 图(a)为 2017 年数据, 图(b)为 2018 年数据, 整体数据见图(c)。

通过对 524 篇基金论文的内容分析发现, 教育学领域学者重用数据目的是新的研究比例最高的种

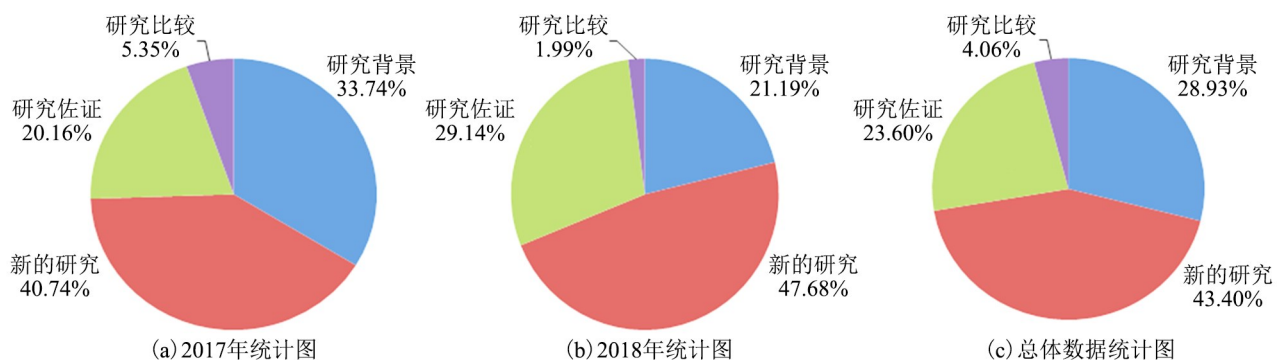


图 1 数据重用目的统计分析图
Fig. 1 Data Reuse Purpose Statistical Analysis

类(43.40%)，其次是研究背景(28.93%)和研究佐证(23.60%)，较低比例的数据重用目的是研究结果比较(4.06%)。与2017年相比，2018年数据重用目的为新的研究及研究佐证的比例上升，在一定程度上可以反映出教育学领域学者越来越注重科学数据价值的挖掘。

2.2 数据重用比例

本文通过标注与统计分析基金论文中的数据重用现象，来获取教育学领域学者数据重用行为的总体情况。2017年立项的49项各类基金项目共产出基金项目论文300篇，其中实证论文篇数为88篇，

占已发表论文总数的29.33%；有数据引用行为的论文有149篇，占比49.67%。2018年立项的51项各类基金项目共产出基金项目论文244篇，其中实证论文篇数为66篇，占已发表论文总数的27.05%，有数据重用行为的论文有109篇，占比44.67%。总体而言，这两年统计结果相差不大，数据重用比例接近50%，这是一个相对比较高的比例。实证论文比例接近30%，如表4所示。由此可见，这两年的数据重用比例都明显高于实证论文比例，即教育学领域学者数据重用比例高于自行收集数据比例。

表 4 数据重用总体情况表
Tab. 4 Data Reuse Overall Situation

时间	基金项目数量	发表论文篇数	数据重用论文篇数	数据重用论文比例 (%)	实证论文篇数	实证论文比例 (%)
2017	49	300	149	49.67	88	29.33
2018	51	244	109	44.67	66	27.05
整体	100	544	258	47.43	154	28.31

调查法在教育学领域适用范围广泛、应用成熟，形成了一定的研究范式，实证论文中调查类论文占比最高。不过这并不意味着教育学领域学者更愿意重用数据而不是自行收集数据，仍有许多学者倾向于通过自行收集数据的方式以实现自己的研究目的。在学者自行收集数据开展实证研究并同时重用他人数据的情况下，数据重用的目的一般为用作研究背景或研究佐证，基本不是用作新的研究。当然，学者也可以使用已有数据，采取思辨方法验证研究论点，或是将原始数据用作实证研究论文中的主要数据和研究基础。因此，教育学领域学者数据重用行为与学者自行收集数据并不矛盾，也没有必然联系。

统计发现，2018年与2017年相比数据重用频次明显下降，一篇论文数据重用次数大于1次的比例下降。2017年数据重用频次为244次，平均每篇论文数据重用次数为1.64次。2018年数据重用频次为151次，平均每篇论文数据重用次数为1.04次。2017年共有5项基金课题项目未见有数据重用现象，占当年抽取比例的10.20%，2018年有11项基金课题项目未见有数据重用现象，占当年抽取

比例的21.57%。可以推测，数据重用行为在一定程度上受数据时间周期以及每项课题发表论文数量的因素影响。

2.3 数据重用程度

由表5可知，2017年未采取任何数据分析手段的数据使用比例为54.73%，2018年数据使用比例为63.58%。

表 5 数据重用处理与分析整理表
Tab. 5 Data Reuse Processing and Analysis

时间	数据处理与分析	
	数据使用比例 (%)	整理成图表比例 (%)
2017	54.73	20.99
2018	63.58	27.15
整体	58.12	23.35

利用原始数据开展新的研究大多需要将原始数据进行再分析，但研究结果显示，教育学领域学者在研究中将原始数据用做新的研究比例不足50%，

而数据分析与处理类目为数据使用的比例超过一半，这表明数据重用目的与数据分析处理方式高度相关，教育学领域学者数据重用目的是研究背景和研究佐证的比例较高，意味着对已有数据的分析不够，对科学数据价值的挖掘不够深入，数据二次分析较少，数据利用程度较低。

2.4 数据重用类型

教育学领域基金论文主要是引用政府网站及期

刊论文中的数据，调查类数据一般来源于官方统计网站以及调查类研究论文，另外有少量数据来自互联网等途径。如表 6 所示，2017 年数据重用类型中，调查数据占比 58.02%，非调查数据占比 41.98%；2018 年数据重用类型中调查数据占比 51.66%，非调查数据占比 48.34%。原始数据类型以调查数据居多，一般是为了引出或验证研究问题，即用做研究背景和佐证观点。

表 6 数据重用类型总体情况表
Tab. 6 Data Reuse Types in General

时间	数据重用 论文篇数	数据重用 次数	篇均数据 重用次数	数据重用类型			
				调查数据	非调查 数据	调查数据 比例 (%)	非调查 数据比例 (%)
2017	149	243	1.63	141	103	58.02	42.39
2018	109	151	1.39	78	73	51.66	48.34
整体	258	394	1.50	219	176	55.58	44.67

注：篇均数据重用次数计算保留小数点后两位。

其中，政府网站及科研院所的数据具有一定的连续性且数据质量较高，如教育部、国家统计局的调查数据涵盖范围广，数据较为权威，可以很好地引出研究背景，是科学研究的重要数据来源。其次是期刊论文和硕博论文，获取较为便捷，并且已发表的通过调查等实证方法收集的数据能够佐证本文观点，也可通过借鉴量表等调查工具开展新的研究。

教育学领域在调查研究中涉及到研究工具的重用，其中量表是比较有特色的数据重用类型。教育学领域量表重用现象较为普遍，可以是借鉴、改编、编译国内外已有研究的成熟量表。通常情况下，对外文量表的数据重用需要翻译审定，并且为契合研究目的应对量表进行本土化及删减，例如“量表经专家使用英文原版翻译……”。对已有成熟的量表的数据重用可以提高研究的科学性，节约时间和精力，是数据重用较为合适的选择。

3 教育学领域基金论文中的数据重用标注规范分析

数据重用来源、格式著录是否规范影响着科学研究的准确性和科学性。因此，数据重用应遵循相应的规范以便科研人员在研究过程中追溯数据来

源，促进科学数据再利用。本文主要对数据重用来源及标注和数据重用选择偏好分析我国教育学领域科研人员数据重用规范。

3.1 数据重用来源及标注

3.1.1 数据重用来源标注

通过统计分析发现，教育学领域学者在数据重用过程中没有任何标注行为(无参考文献、无注释、无文中提及)的比例在 10%以下，如表 7 所示。

表 7 数据重用规范总体情况表

Tab. 7 Data Reuse Specifications

时间	数据重 用频次	数据来 源未知	数据来 源未知 比例 (%)	数据重 用规范 “无” 比例 (%)	数据重 用规范 “无” 比例 (%)
2017	243	16	6.58	13	5.35
2018	151	25	16.56	15	9.93
整体	394	41	10.41	28	7.11

2017 年“无”的比例仅为 5%，2018 年为 9.93%，即超过 90%的学者会在数据重用中标注原始数据

来源, 这表明绝大部分学者具备一定的数据重用规范意识。但是在数据来源方面, 2017 年数据来源未知比例有 6.58%, 2018 年数据来源未知比例有 16.56%。整体而言, 有 10% 左右的学者没有明确指出原始数据来源渠道, 均大于数据重用规范为“无”的比例。这说明即便学者在文中提及数据来源或进行数据引用, 但未必能全部明确指出详细完整的数据来源渠道。

教育学领域学者重用数据存在不介绍重用数据的来源, 或者交待部分出处但不可查的问题, 表现为对原始数据不作实质性出处说明, 不标注引用数据以及对于数据记录的完整性认识不足等形式。数据重用更注重数据的结果, 较少关注能够反映数据全过程信息的元数据。对于科研人员自己整理的数

3.1.2 数据重用标注格式

教育学领域学者在数据重用过程中很少像自然科学领域对数据或数据集进行明确声明, 对原始数据的来源标注一般通过文中提及、参考文献及注释的方法, 其中参考文献形式是科学数据的正式引用形式, 其他 3 类是非正式形式。但科研人员标注数据来源时, 往往不只采用一种形式, 同时使用文中提及和参考文献的情况较多, 以保证数据的准确性和权威性。

统计发现, 教育学领域数据重用格式为注释这一形式较少, 2017 年仅 16 条数据使用注释这一形式, 2018 年仅 3 条数据通过注释说明原始数据来源。注释可以详细介绍数据的获取及相关说明, 但较少采用注释方式注明数据来源。涉及到对历史资料的数据重用, 注释这一非正式引用方式的较多, 如注释中写道“此表内容来源于……的归纳”。注释一般提及创建者及标题两项, 实际上, 注释与参考文献作用相同, 存在形式上是注释, 实际是参考文献的情况, 但通常注释与参考文献标注的格式不一致。

3.1.3 数据重用标注习惯

高校及研究院等权威机构的调查统计报告也是数据重用的重要来源渠道, 但与此相关的数据重用

规范却并不统一。若是仅对调查统计结果进行数据重用, 则更倾向于在论文正文标注参考文献。如果一开始就没有将调查数据列入参考文献中, 那之后也不会列入, 这是一种论文写作习惯。一般直接使用调查数据, 在正文中不会说明数据的增删改情况。但实际情况复杂得多, 科学数据类型多样, 除调查数据之外, 教育学领域学者重用非调查数据的情况较为常见, 测量量表和评价指标是出现频率较高的非调查数据。当科研人员重用这类数据时一般会以往研究整理成表格形式, 部分以参考文献的方式出现, 参考文献如果是期刊论文则较为规范, 一般按照期刊论文的著录标准著录。如果来源于网站, 那么在表格后表注释数据来源则更为常见。而此种方式数据引用形式最为多样, 表明教育学领域学者数据规范意识不足。

通过对数据重用来源及标注的分析可知, 虽然我国已发布科学数据相关国家标准, 但在实践中仍存在大量不规范数据重用现象, 主要体现为科研人员在数据重用过程中未明确注明数据重用来源, 数据规范意识有待提高。

3.2 数据重用选择偏好

3.2.1 重用位置特征

为了解教育学领域学者数据重用行为规范, 本研究进一步分析数据引用位置, 将数据重用位置分为引言、正文和结论 3 部分。一般而言, 在引言部分数据重用目的为研究背景, 但同时也有少部分作为背景资料出现在正文中。总体而言, 在正文中有数据重用的比例最大, 因为正文内容是研究的主体部分, 且在正文中数据重用目的可以是用作研究背景、开展新的研究以及佐证研究观点。教育学领域数据重用位置主要集中在引言和正文, 以科学数据为研究基础的论文通常会在正文研究设计中提到数据来源。数据重用出现在结论的部分占比最少, 在论文结论部分进行数据重用一般是为了研究结果比较分析, 而数据重用为研究比较的比例较低。因此, 可以发现教育学领域学者较为重视科学数据的价值。

3.2.2 数据重用元素选择

科学数据元素是科学数据的进一步揭示, 能够反映数据重用行为是否规范。不同数据类型的元数据元素差异较大, 即便是相同数据类型在同期刊

也不一样,存在信息来源不完整、欠精确,对来源信息揭示度不高的问题。目前数据引用元素的规范并不统一,必备元素和可选元素的标准存在较大差异。为了能够对数据集进行溯源,需要统一资源标识符。

当前数据重用不规范情况大量存在,虽然已经说明了数据来源,但因为缺乏数据重用规范和数据应用格式的认识,存在数据重用格式不规范、信息著录不完善、数据来源标注较为随意等问题,导致论文作者和原始数据查找困难。当前学者、期刊和相关机构缺乏对数据引用规范性的认识,使论文的数据引用不规范问题普遍存在。学者在论文中虽然重用了相关数据,但并没有对这些数据及其来源明确说明和注释。数据的引用格式也存在问题,使数据信息展示不全,甚至没有详细标注引用数据及其来源。

在数据重用过程中,数据元素标注还不够全面,存在数据来源后只有原始数据的标题和时间信息的现象。大多数学者在引用图表形式的数据时,仅在文中描述数据来源,列出部分数据元素,此类不规范的数据重用行为影响了科研人员通过数据引用信息定位原始数据,阻碍了数据的传播与再利用。

3.2.3 数据重用行为相似性

马太效应反映了“强者越强,弱者越弱”的现象,教育学领域科研人员数据重用行为存在“马太效应”,这在数据重用行为中表现为一项基金项目在项目前期重用他人研究数据,之后数据重用的可能性大为提高。反之,如果一项基金项目在项目前期没有重用他人研究数据,项目中后期很可能也不会有数据重用行为。2017 年共有 5 项基金课题项目未见有数据重用现象,占当年抽取比例的 10.20%,2018 年有 11 项基金课题项目未见有数据重用现象,占当年抽取比例的 21.57%。这两年的数据体现了“马太效应”中“弱者越弱”的现象。与此对应的是 2017 年共有 7 项基金课题项目数据重用论文比例为 100%,即已发表的每一篇论文都有数据重用行为,2018 年数据重用论文比例为 100% 的基金课题有 11 项,但其中有 9 项为仅发表了 1 篇论文,并且这两年数据重用比例为 100% 的基金课题项目已发表论文数均不超过 5 篇,由此可知数据重

用并不是教育学领域科研人员开展研究的首选。

通过分析发表论文数大于 5 篇的基金论文中的数据重用行为发现,系列论文的数据重用行为特征也具有相似性。由于系列论文具有相对一致的研究方向,同一项目数据重用类型存在一致性倾向,倾向于使用同一类型数据的情况,这与研究内容、研究方法、数据重用目的以及研究者的个人习惯有关。

4 结 语

本研究采用内容分析法从相对客观的角度分析教育学领域数据重用行为特征规律,选取了全国教育科学规划项目数据为总体数据,对 2017 年和 2018 年的全国教育科学规划立项课题进行抽样,从数据重用目的、数据类型偏好及数据重用规范等方面分析教育学领域科研人员数据重用行为。研究发现,教育学领域科研人员数据重用行为有以下特征:

①教育学领域科研人员数据重用比例较高且较为稳定;②教育学领域科研人员偏向于调查数据的重用,其他数据以公式、代码和教学数据资源为主,对量表的编译改编借鉴重用较为频繁;③在数据来源上,政府网站及数据中心是科研人员较为信任的数据来源渠道,已发表的专业期刊论文中的科学数据由于数据获取便利性和专业性成为科研人员频繁引用的原始数据。此外,互联网数据因其可获得性和时效性较强得到科研人员的关注;④科学数据重用过程中的数据分析与处理较为粗浅,数据使用最多,将原始数据整理成表格也是常见的数据处理分析方式,文献计量与数据可视化分析逐步成为热门,但存在数据分析方法较为粗浅的问题。并且数据分析处理与数据重用目的密切相关;⑤由于缺少明确统一的科学数据参考文献及注释格式规范,科学数据的元数据描述差异较大。此外,数据重用过程数据来源不详是数据获取障碍的重要原因。

本研究以教育学领域的基金项目产出论文为分析依据,对其中的数据重用现象进行了识别和提取,通过对该领域学者数据重用行为特征的分析,总结其专业内部的科学数据使用偏好及使用规范,而数据重用行为有明显的学科差异性。本研究通过对教育学领域科研人员数据重用行为的分析,希望能够在一定程度上为图书馆、数据出版商、数据中心等机构开展数据服务内容及调查数据平台等教学资源

数据库的建设方向提供参考。

但本研究仅从相对客观的角度分析科研人员数据重用行为,关注了基金论文文本中体现的数据重用现象,未能探讨文本外的科研人员自身的情感、态度等主观方面的动态的因素。在基金项目样本的时间范围的选择上,本次研究仅选择了 2017 年和 2018 这两年的基金项目,相对而言数据量有一定局限性。另外,数据重用规范仅仅考虑到格式、位置及元素,更偏重研究数据重用行为形式特征,内容分析编码类目构建考虑不够完善,未来研究还需挖掘数据重用行为其他特征表现。

参 考 文 献

- [1] Szabo V, Strang V R. Secondary Analysis of Qualitative Data [J]. *Advances in Nursing Science*, 1997, 20 (2): 66-74.
- [2] Castle J E. Maximizing Research Opportunities: Secondary Data Analysis [J]. *Journal of Neuroscience Nursing*, 2003, 35 (5): 287-290.
- [3] Zimmerman A S. New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data [J]. *Science, Technology & Human Values*, 2008, 33 (5): 631-652.
- [4] Yoon A. Data Reusers' Trust Development [J]. *Journal of the Association for Information Science and Technology*, 2017, 68 (4): 946-956.
- [5] Pasquetto I V, Randles B M, Borgman C L. On the Reuse of Scientific Data [J]. *Data Science Journal*, 2017, 16 (1): 1-9.
- [6] Custers B H M, Helena U V. 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, (1): 4-15.
- [7] Federer L. Who, What, When, Where and Why? Quantifying and Understanding Biomedical Data Reuse [EB/OL]. <http://hdl.handle.net/1903/21991>, 2021-12-01.
- [8] Youngseek K. A Study of the Roles of Metadata Standard and Data Repository in Science, Technology, Engineering and Mathematics Researchers' Data Reuse [J]. *Online Information Review*, 2021, 45 (7): 1306-1321.
- [9] Park M S, Park H. An Examination of Metadata Practices for Research Data Reuse: Characteristics and Predictive Probability of Metadata Elements [J]. *Malaysian Journal of Library & Information Science*, 2019, 24 (3): 61-75.
- [10] Faniel I M, Kriesberg A, Yakel E. Social Scientists' Satisfaction with Data Reuse [J]. *Journal of the Association for Information Science and Technology*, 2016, 67 (6): 1404-1416.
- [11] Joo Y K, Youngseek K. Engineering Researchers' Data Reuse Behaviours: A Structural Equation Modelling Approach [J]. *Electronic Library*, 2017, 35 (6): 1141-1161.
- [12] Curty R G. Beyond "Data Thrifting": An Investigation of Factors Influencing Research Data Reuse in the Social Sciences [D]. Univ. Syracuse, 2015.
- [13] Yoon A. Data Reusers' Trust Development [J]. *Journal of the Association for Information Science & Technology*, 2017: 68 (4): 946-956.
- [14] Tenopir C, Dalton E D, Allard S, et al. Changes in Data Sharing and Data Reuse Practices and Perceptions Among Scientists Worldwide [J/OL]. <https://doi.org/10.1371/journal.pone.0134826>, 2021-12-01.
- [15] Yoon A, Youngseek K. The Role of Data-reuse Experience in Biological Scientists' Data Sharing: An Empirical Analysis [J]. *Electronic Library*, 2020, 38 (1): 186-208.
- [16] Joon S, Youngseek K. 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.
- [17] Yoon A, Youngseek K. 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.
- [18] Bishop L, Kuula L A. Revisiting Qualitative Data Reuse: A Decade On [J/OL]. doi:10.1177/2158244016685136, 2021-12-01.
- [19] Youngseek K, Yoon A. Scientists' Data Reuse Behaviors: A Multilevel Analysis [J]. *Journal of the Association for Information Science and Technology*, 2017, 68 (12): 2709-2719.
- [20] 陈雅芝. 欧盟地理信息再利用的有效实践与启示 [J]. *情报资料工作*, 2011, (6): 60-65.
- [21] 赵蕊菡. 科学数据论文的重用现状研究——基于数据期刊“Earth System Science Data”的引文分析 [J]. *情报理论与实践*, 2017, 40 (11): 52-57, 72.
- [22] 林奇秀, 赖璟毅. 台湾社会科学量化资料再用之研究: 2001—2015 年 [J]. *教育资料与图书馆学*, 2018, 55 (1): 39-69.
- [23] 孙玉伟, 成颖, 谢娟. 科研人员数据复用行为研究: 系统综述与元综合 [J]. *中国图书馆学报*, 2019, 45 (3): 110-130.
- [24] 张莹, 戚景琳, 孙玉伟. 管理学科科研人员数据复用行为特征探析 [J]. *信息资源管理学报*, 2020, 10 (4): 79-87.
- [25] 戚景琳, 张莹, 孙玉伟. 社会科学科研人员数据复用行为研究——以经济学为例 [J]. *情报理论与实践*, 2020, 43 (9): 72-78.
- [26] 焦红, 杨波, 周琪. 生物医学领域科学数据集复用特征研究 [J]. *情报理论与实践*, 2021, 44 (9): 90-96.
- [27] 王雪, 杨波. 科学数据重复使用的学科差异性研究 [J]. *情报杂志*, 2021, 40 (7): 122-126, 156.

- [28] 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]. 长春: 东北师范大学, 2019.
- [29] 张潇月, 宋秀芬, 顾立平, 等. 我国科研人员科研数据重用行为影响因素实证研究——以生物学领域为例 [J]. 情报学报, 2021, 40 (8): 887-902.
- [30] 孙玉伟, 戚景琳, 刘超, 等. 社会科学科研人员数据复用行为影响因素研究: MOA 框架下的探索 [J]. 情报理论与实践, 2021, 44 (10): 90-99, 64.
- [31] 文静, 何琳, 韩正彪. 科研人员科学数据重用意愿的影响因素研究 [J]. 图书情报知识, 2019, (1): 11-20.
- [32] 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): 11-16.
- [33] 张彤彤. 科研数据重用行为及影响因素研究 [D]. 长春: 吉林大学, 2020.
- [34] 文静. 科研人员科学数据重用意愿影响因素研究 [D]. 南京: 南京农业大学, 2018.
- [35] 张若晗. 科研人员数据重用满意度影响因素研究 [D]. 哈尔滨: 黑龙江大学, 2021.
- [36] 孙玉伟. 基于“方差”和“过程”理论的科研人员数据复用行为研究逻辑框架 [J]. 图书馆学研究, 2020, (18): 70-79.
- [37] 孟琰. 开放科学数据平台的数据复用影响因素研究 [D]. 上海: 华东师范大学, 2022.
- [38] 宋秀芬, 周茜, 李立睿, 等. 科研数据再利用过程模型研究 [J]. 情报理论与实践, 2021, 44 (6): 61-68.
- [39] 宋秀芬, 周茜, 李立睿, 等. 科研数据再利用性的五维度模型研究 [J]. 现代情报, 2021, 41 (10): 83-92.
- [40] Wang X G, Duan Q Y, Liang M L. Understanding the Process of Data Reuse: An Extensive Review [J]. Journal of the Association for Information Science and Technology, 2021, 72 (9): 1161-1182.
- [41] 黄欣卓, 米加宁, 章昌平, 等. 科学数据复用研究的演化、知识体系与方法工具——兼论第四科研范式的影响 [J]. 科研管理, 2022, 43 (8): 100-108.
- [42] 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, (3): 80-89.
- [43] 张闪闪, 刘晶晶, 顾立平, 等. 科研数据内容重用中的权益问题研究 [J]. 图书情报知识, 2018, (1): 105-113, 94.
- [44] 张潇月, 宋秀芬, 顾立平, 等. 我国科研人员科研数据重用行为影响因素实证研究——以生物学领域为例 [J]. 情报学报, 2021, 40 (8): 887-902.
- [45] 邱均平, 何文静. 科学数据共享与引用行为的相互作用关系研究 [J]. 情报理论与实践, 2015, 38 (10): 1-5.
- [46] 史雅莉, 司莉. 我国科研人员数据引用行为特征分析 [J]. 情报理论与实践, 2019, 42 (6): 36-41.
- [47] 刘祥洪. 我国人口学领域数据引用评价研究 [D]. 杭州: 浙江大学, 2016.

(责任编辑: 杨丰侨)

(上接第 41 页)

- [31] 李海波, 高晓瑾, 丁华, 等. 科技发展决策支持平台运行机制的体系构成及要素分析 [J]. 科技与管理, 2012, 14 (2): 28-31.
- [32] 梁永霞, 杨中楷. 科技期刊编辑应增强科学学素养 [J]. 中国科技期刊研究, 2020, 31 (4): 425-431.
- [33] 卢泰宏. 论情报和情报系统的作用机理 [J]. 情报科学, 1982, 3 (2): 15-21.
- [34] 张云龙, 马淑欣. 论科技发展与人文精神的内在勾连——基于世界科学中心转移的视角 [J]. 自然辩证法研究, 2022, 38 (2): 86-92.
- [35] 陈悦, 宋超, 徐芳. 我国“科学—技术—经济”产出的动态关系测度研究——基于“科学—技术”交互的视角 [J]. 科研管理, 2019, 40 (1): 12-21.
- [36] 朱铁壁, 张红霞. 高校分类新思考: 知识生产与学生学习双重视角 [J]. 高等教育研究, 2015, 36 (11): 24-30.
- [37] 姜春林. 大学知识能力及其测度指标体系初探 [J]. 辽宁工程技术大学学报 (社会科学版), 2007, 46 (6): 612-614.
- [38] 宁烨, 樊治平. 知识能力的构成要素: 一个实证研究 [J]. 管理评论, 2010, 22 (12): 96-103.
- [39] 宁凌, 苏玉同, 欧春尧. 三螺旋理论视角下区域海洋高新技术产业创新效率分析与评价研究 [J]. 科技管理研究, 2022, 42 (15): 57-64.
- [40] 魏春艳, 方益权, 衡孝庆. 基于知识形态的新工科产教融合机理探究 [J]. 中国高教研究, 2022, 342 (2): 89-94.
- [41] 齐燕. 基于专利引文分析的领域科技关联发展态势探究——以 HCV 为例 [J]. 图书情报工作, 2019, 63 (23): 97-105.
- [42] 陈悦, 刘则渊. 悄然兴起的科学知识图谱 [J]. 科学学研究, 2005, (2): 149-154.
- [43] 梁永霞, 杨中楷, 王贤文. 刘则渊与中国知识计量学 [J]. 图书情报工作, 2013, 57 (11): 137-142.
- [44] 毕崇武, 叶光辉, 夏立新. 引文内容分析视角下图书情报领域知识计量研究刍议 [J]. 图书情报工作, 2021, 65 (21): 97-107.
- [45] 孙晓玲, 丁堃. 深度学习中的表示学习研究及其对知识计量的影响 [J]. 情报理论与实践, 2018, 41 (9): 118-122.
- [46] 孙震, 冷伏海. 基于知识元的新型科学计量范式探析 [J]. 情报学报, 2017, 36 (6): 555-564.

(责任编辑: 杨丰侨)