

我国图书情报领域数据引用行为分析^{*}

丁楠 丁莹 杨柳 凌晨 潘有能

摘要 本文以内容分析法为主要研究方法,对我国图书情报领域权威期刊2003年至2013年间的数据库引用行为进行调查分析。研究发现:数据库引用在图书情报研究领域中的重要性正在逐步提升,总体而言虽然规范引用比率呈现逐渐走高的状态,但非规范引用比率仍然不可小视;《中国图书馆学报》和《大学图书馆学报》的数据库引用现象较《情报学报》要更多一些,但《情报学报》的数据库引用相对更为规范;数据库引用行为较少及不规范引用大量存在的主要原因在于大量数据没有得到有效应用,已有数据可获得性较差,以及缺乏规范的数据引用格式。图2。表6。参考文献9。

关键词 图书情报学 数据库引用 学术规范

分类号 G350

Data Citation Behavior in Library and Information Science in China

Ding Nan, Ding Ying, Yang Liu, Ling Chen & Pan Youneng

ABSTRACT Using the method of content analysis, the authors studied the data citation behavior of the papers from top journals in library and information science in China between 2003 and 2013. The results show: Data citation is becoming more and more important in the field of library and information science. Although the percentage of standard citation is increasing gradually, the percentage of non-standard citation can not be ignored; The data citation behavior in the papers from *Journal of Library Science in China* and *Journal of Academic Libraries* is more frequent than in the papers from *Journal of the China Society for Scientific and Technical Information*, but the data citation behavior in the papers from *Journal of the China Society for Scientific and Technical Information* is more standardized; The main reasons of the lack and non-standard of data citation behavior include the lack of effective use of mass data, the poor accessibility of the existing data, and the lack of standard format of data citation. 2 figs. 6 tabs. 9 refs.

KEY WORDS Library and information science. Data citation. Academic norm.

0 引言

随着现代科学的发展,数据已成为科学研究不可或缺的部分,在已有数据的基础上开展研究是科学研究的基本模式之一。近年来,数据管理与共享得到了越来越多的关注^[1-2],但由于激励

和评价机制的缺失,以及技术和规范的不成熟,机构和作者的数据共享行为显得前途不明且动力不足,影响了数据共享的进一步推广。

数据库引用是指作者在文献中以参考文献、脚注或文中注的方式,对其所引用数据提供来源出处的做法。可以看出,数据库引用和目前已在图书情报领域得到深入研究的文献引用具有

^{*} 本文系浙江省哲学社会科学规划项目“社会科学数据库引用行为与评价机制研究”(编号:13NDJC130YB)的研究成果之一。

通讯作者:丁楠,Email:nding@zju.edu.cn

较高的相似度。文献引用以及在其基础上发展起来的引文分析已广泛应用到文献检索、科研管理、学术评价等领域,与之相类似,数据引用也可在数据的检索、管理和评价等方面发挥重要作用,具体体现为:

(1) 数据引用一般应包括数据的作者、标题、出处以及访问地址等信息,读者可利用这些信息定位并获取数据,从而促进数据共享与重用;

(2) 数据引用可增加数据生产者 and 数据存储机构的被认可度,成为评价这些人员和机构贡献的重要依据,从而对数据共享起到一定的激励作用;

(3) 通过数据引用可以追踪数据的使用,降低数据被剽窃的风险;

(4) 通过数据引用可获取科学研究的原始数据,按照研究过程实施,可重现研究过程,从而验证科学研究的结果^[3]。

国外图书情报领域的研究学者很早就意识到了数据引用的重要性,著名的科学计量学家、霍普斯奖获得者 Howard D. White 早在 20 世纪 80 年代便指出社会科学学者应该像引用文献一样以规范化的格式列出其使用的数据^[4],但未能得到学界和出版界的充分重视。Valerie 选择了 6 种期刊在 2000—2010 年间发表的 500 篇文献作为样本,研究后发现只有约 14% 的文献标出了所使用数据集的唯一标识,另外有约 12% 的文献提到了数据集作者和数据仓储的名称,但著录不够完整,总体来看很少有文献将其对数据集的引用纳入参考文献中^[5]。Hailey 和 Mark 通过检索 WilsonWeb 数据库中文摘含有关于数据或数据集的存储、检索、利用、分析等相关词汇的论文,并在检索结果中按人文科学、社会科学和自然科学三个类别随机抽样,共选取了 65 篇论文,这些论文分属于 44 种期刊。分析结果显示,大约一半的期刊刊发关于数据引用方面的声明及格式要求,但对这些期刊上的论文进行内容分析后发现,大部分论文在数据引用方面有较大的缺陷^[6]。Hailey 和 Mark 的研究结论具有一定的参考价值,但其样本数量过少,

且分属于各个学科,因此无法针对某个学科或期刊进行深入探讨。目前尚未见到针对图书情报领域数据引用行为的研究。

在国内,关于数据引用情况调查的论文尚不多见,墨愚曾对经济类论文进行过一次统计分析,选取 2001 至 2003 年间《新华文摘》上 60 篇有数据引用行为的经济类论文为研究对象,发现这些论文所引用的 281 个(组)数据中,78.15% 的数据没有任何实质性出处说明,由此可见数据引用的问题较为严重^[7]。值得注意的是,墨愚的研究以数据组为单位,而本文的研究以论文为单位。这可能与墨愚关注点在于学术规范,而本文的着眼点在于考查学科数据引用特征有关。

本研究拟对我国图书情报学领域数据引用行为进行抽样调查和研究,具有三方面意义。

(1) 加深对图书情报学学科的了解和认识。在该学科以往的定量研究中,学者对文献的各种著录元素及文献之间的引用关系都已研究得相当深入和详尽,但是对数据引用的研究则十分薄弱,尤其是对该学科数据引用行为的研究更是尚未见到。本研究有助于了解图书情报学中科研论文对数据的倚赖程度,并进一步揭示学科特征,完善学科认识。

(2) 科学评价图书情报学的数据引用现状。著录和引用是图书情报学研究的重要组成部分,因此从学科性质和研究内容的角度而言,图书情报学也理应在数据引用规范上更加严谨,成为其他学科的表率。然而其引用现状究竟如何?通过探讨该学科引用行为的特点、引用规范现状及可能的原因,有望给出合理的解释并有针对性地提出解决策略。

(3) 对数据引用行为的分析可为领域内科研人员、管理部门及出版机构提供借鉴。可以预见的是引用行为的规范化不但可以促进学术规范,规避学术不端,还能促进数据的共享与交流。

1 数据与方法

本研究采用等距抽样方法,选取国内图书

情报领域的权威期刊《中国图书馆学报》(以下简称《中图学报》)、《情报学报》和《大学图书馆学报》(以下简称《大图学报》)2003—2013 年共 11 年每年上半年第 1 期和下半年第 1 期(即双月刊每年的 1、4 两期,单月刊每年的 1、7 两期,共计 66 册期刊)刊登的论文为统计对象,除去编辑寄语、会议报导、简讯、回顾性散文等,共计 1 290 篇论文。本研究所选三种刊物均为图书情报领域的权威期刊,在某种程度上代表了领域内科研的最高水平,同时每期发表文章篇数也较为接近,因此具有代表性和可比性。

本研究采用内容分析法对上述 1 290 篇样本论文进行数据引用行为的识别。由于数据引用还没有规范的格式及可靠的机器识别方法,本研究主要采用人工识别的方式,因此建立明确、细致的数据处理规范十分重要。为减少标引员主观判断的影响,在标引前对标引员进行了多轮培训,并选取样本材料,由两个标引员分别进行标引,结果显示其信度系数达到了 95% 以上,保证了分析结果的可靠性和客观性。

数字的出现是数据引用最基本、最明显的特征,但是数据引用的识别远非数字识别那么简单,除了区分公式等由数字组成但并非数据引用的明显例子,还需对各种复杂情况进行仔细判别。

(1)需区分数据引用与事实陈述。在进行事实陈述尤其是背景介绍时,往往会出现数字,如“十余年来……”、“1994 年……”等,此时虽然有数字出现,但是并非数据引用的范畴。而有时,数据引用与事实陈述的界限会比较模糊,如“其发展高峰期,主要出现于…年到…年”,此时,数据作为表达的中心和主体,可认为其属于数据引用范畴。

(2)需区分数据引用与数据提供。有时,文献著者也会提供一些数据作为研究成果,此时,只需仔细阅读就可与数据引用相区别。而采用文献计量学作为主要研究方法的著者,经常会从数据库中检索并获取大量数据进行分析,这也不同于数据引用。

基于当前数据引用在整体上呈现出的复杂状况,在判定数据引用的基础上,需要再对引用类型作进一步划分。类似于文献引用,数据引用一般来说也应包括作者、数据标题、出版机构、出版时间、访问地址等要素。数据引用可以有两种划分方式:有引用和无引用,规范引用和不规范引用。也可以进一步细分为规范引用、部分引用和无引用。如表 1 所示。

表 1 数据引用行为分类表

分类 1	分类 2	分类 3
有引用	规范引用	规范引用
	部分引用	不规范引用
无引用	无引用	

数据引用的划分,主要根据论文中引用要素的完整程度。规范引用所包含要素是否完整,其重要特征是根据要素找到原始数据,目前较为常见的有表后注和参考文献引用。部分引用往往列出了数据的部分信息,如仅提到数据作者、机构的名称或出处,很难据此找到原始数据,而只是知道大致的数据来源。如“由麦肯锡公司完成的一项研究推断……”,提供了数据发布的机构名称,但是数据的详细出处仍然无从获悉。而无引用则直接罗列数据,完全不交代数据来源。“据统计……”、“据调查……”和“各种研究表明……”是无引用的常见形式,如:“据统计,竞争情报分析所需的 90% 的信息可以通过 Internet 进行收集和分析。”

在实际操作中,无引用相对容易判断,但是规范引用和部分引用的界限有时比较模糊。此时,是否可以根据标引信息找到数据源,以及是否存在进一步标引的可能,可以作为判断依据。而这种情况的存在,也客观上说明了建立数据引用规范的必要性和迫切性。

在引用次数上,也存在一些需要特别说明的情况:一个数据来源可能在论文中被多处引用,则每一处引用都计入引用次数;一处数据可能同时出现数个标引,则根据标引个数统计其引用次数。

2 结果与讨论

2.1 总体数据及分析

经数据处理及汇总,图书情报领域数据引用的基本情况如表2所示。

表2反映了2003—2013年11年间,图书情报领域三个权威期刊的数据引用情况。从中可以看出,2003—2008年间的篇均数据引用次数均在1~2次之间,从2009年开始有了较大提升,2013年达到最高值2.7次,说明数据引用在图书情报研究领域中的重要性正在逐步提升。

表2 图书情报领域数据引用情况表

年份	总论文数	总引用次数	规范引用次数	部分引用次数	无引用次数	篇均引用次数	规范引用比率(%)	部分引用比率(%)	无引用比率(%)
2003	133	250	116	17	117	1.88	46.40	6.80	46.80
2004	128	168	61	41	56	1.31	36.31	24.40	33.33
2005	122	228	133	27	68	1.87	58.33	11.84	29.82
2006	124	152	88	15	49	1.23	57.89	9.87	32.24
2007	133	254	177	13	64	1.91	69.69	5.12	25.20
2008	128	140	96	13	29	1.09	68.57	9.29	20.71
2009	114	241	186	10	45	2.11	77.18	4.15	18.67
2010	122	321	256	17	48	2.63	79.75	5.30	14.95
2011	98	206	162	8	36	2.10	78.64	3.88	17.48
2012	94	202	170	9	23	2.15	84.16	4.46	11.39
2013	94	254	169	45	40	2.70	66.54	17.72	15.75
总计	1 290	2 416	1 614	215	575	1.87	66.80	8.90	23.80

从数据引用行为的年度分布来看,11年间,仅有2003年和2004年的规范引用比率低于50%,大部分年份的规范引用比率介于50%到80%之间,2012年的规范引用比率最高,超过80%。2003—2013年平均规范引用比率约为67%,也即2/3的数据引用均为规范引用。其中

前半段,即2003—2008年的规范引用比率相对较低,后半段,即2009—2013年间除2013年外均高于70%。虽然2013年的规范引用比率有所下降,但总体来看规范引用比率仍呈现出逐渐走高的态势(见图1)。

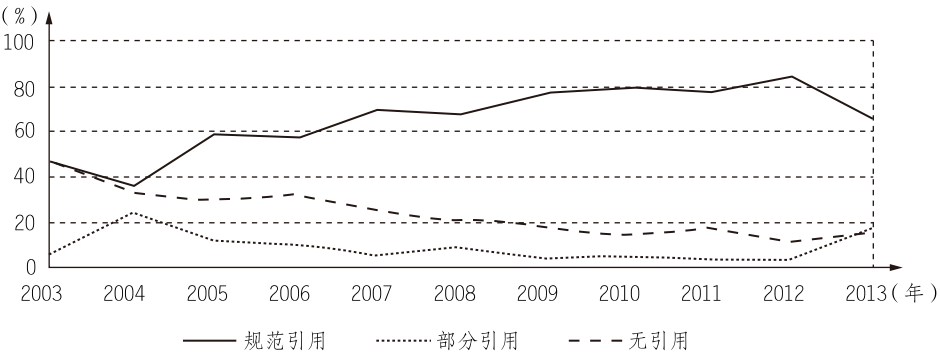


图1 图书情报领域数据引用行为趋势图

此外,由表 2 可知,11 年间三种期刊的部分引用比率仅为 8.9%,远低于规范引用比率。而无引用比率相对较高,2003—2008 年间均超过 20%,2009—2013 年间均介于 10%~20%之间。总体而言,虽然规范引用比率呈现逐渐走高的状态,但非规范引用比率仍然不可小视。

2.2 期刊比较分析

由表 3 可知,2003 到 2013 年间,《中图学报》每期的论文刊载量呈现出逐渐下降的趋势。而篇均数据引用次数在 2009 年之后呈现出明显增加的趋势。在《中图学报》论文的数据引用行为中,部分引用所占比率非常低。2009 年之前,

规范引用和不规范引用的比率非常接近,基本各占 50%左右,但 2009 年起,规范引用的次数远远超过了不规范引用,只有 2013 年出现了一个小的跌宕,这主要是由该刊当年第 1 期的论文《数字化转变:电子书和电子内容对读者和图书馆的影响》造成的,该文以时任美国图书馆协会主席莫莉·拉斐尔访华期间的演讲稿为基础形成,文后无任何参考文献,而文中 10 处数据引用亦无任何来源交代,因此被归类为无引用行为,大大影响了统计结果。

因此,可以认为 2009 年是《中图学报》数据引用的一个转折年,在此之后,该刊单篇论文的数据引用量增大,规范程度也有显著提高。

表 3 《中国图书馆学报》数据引用情况表

年份	总论文数	总引用次数	规范引用次数	部分引用次数	无引用次数	篇均引用次数	规范引用比率 (%)	部分引用比率 (%)	无引用比率 (%)
2003	48	96	38	8	50	2.00	39.58	8.33	52.08
2004	46	89	31	22	26	1.93	34.83	24.72	29.21
2005	47	65	38	1	26	1.38	58.46	1.54	40.00
2006	47	70	39	4	27	1.49	55.71	5.71	38.57
2007	43	92	60	4	28	2.14	65.22	4.35	30.43
2008	40	51	28	8	13	1.28	54.90	15.69	25.49
2009	29	114	102	1	11	3.93	89.47	0.88	9.65
2010	30	89	76	5	8	2.97	85.39	5.62	8.99
2011	26	64	56	2	6	2.46	87.50	3.13	9.38
2012	27	67	64	1	2	2.48	95.52	1.49	2.99
2013	26	80	57	1	22	3.08	71.25	1.25	27.50
总计	409	877	589	57	219	2.14	67.16	6.50	24.97

《大图学报》(见表 4)11 年来的论文刊载量十分稳定,基本保持在每期 20 篇左右。篇均数据引用次数则在 2006—2009 年出现了一个波谷,但从 2010 年起有了较大提高,这也和近年来刊登了较多的史料性和综述性论文有关。《大图学报》的规范引用比率在 2007 年之前为 50%左右,其后则提升为 70%左右,但仍待进一步提高。

《情报学报》(见表 5)每期刊载的论文数总

体呈下降趋势,特别是自 2011 年从双月刊改为月刊之后,每期论文数稳定在 12~15 篇左右,同时单篇论文页数增多、篇幅变长。前文介绍的《中图学报》自 2009 年以来也出现了类似的变化,这从一个侧面反映出我国图书情报领域顶级期刊近年来刊载的论文质量有较大提升。近年来,《情报学报》上几乎没有刊登只有三四页的短篇幅论文,大量的数据以图、表、模型、公式等各种形式拉长了论文的篇幅。然而与数据引

证比较起来,尽管重视数据,学者似乎更热衷于提供数据而非引用数据,因此,即便是大量数据扑面而来的《情报学报》,其篇均数据引用次数并不高,甚至低于《中图学报》和《大图学报》。这也在一定程度上说明了尽管每年产生了大量

的科研数据,但是这些数据并没有得到充分利用。当然,技术性论文占比较高也是《情报学报》篇均数据引用次数较低的原因之一。2008年以后,规范引用的比率同样显著提升,与《中图学报》、《大图学报》十分类似。

表 4 《大学图书馆学报》数据引用情况表

年份	总论文数	总引用次数	规范引用次数	部分引用次数	无引用次数	篇均引用次数	规范引用比率 (%)	部分引用比率 (%)	无引用比率 (%)
2003	42	88	44	4	40	2.10	50.00	4.55	45.45
2004	36	37	18	5	14	1.03	48.65	13.51	37.84
2005	37	93	43	22	28	2.51	46.24	23.66	30.11
2006	40	47	25	8	14	1.18	53.19	17.02	29.79
2007	46	72	52	2	18	1.57	72.22	2.78	25.00
2008	44	58	42	5	11	1.32	72.41	8.62	18.97
2009	42	83	46	6	31	1.98	55.42	7.23	37.35
2010	42	154	108	9	37	3.67	70.13	5.84	24.03
2011	45	117	85	5	27	2.60	72.65	4.27	23.08
2012	42	118	90	8	20	2.81	76.27	6.78	16.95
2013	44	131	75	41	15	2.98	57.25	31.30	11.45
总计	460	998	628	115	255	2.17	62.93	11.52	25.55

表 5 《情报学报》数据引用情况表

年份	总论文数	总引用次数	规范引用次数	部分引用次数	无引用次数	篇均引用次数	规范引用比率 (%)	部分引用比率 (%)	无引用比率 (%)
2003	43	66	34	5	27	1.53	51.52	7.58	40.91
2004	46	42	12	14	16	0.91	28.57	33.33	38.10
2005	38	70	52	4	14	1.84	74.29	5.71	20.00
2006	37	35	24	3	8	0.95	68.57	8.57	22.86
2007	44	90	65	7	18	2.05	72.22	7.78	20.00
2008	44	31	26	0	5	0.70	83.87	0.00	16.13
2009	43	44	38	3	3	1.02	86.36	6.82	6.82
2010	50	78	72	3	3	1.56	92.31	3.85	3.85
2011	27	25	21	1	3	0.93	84.00	4.00	12.00
2012	25	17	16	0	1	0.68	94.12	0.00	5.88
2013	24	43	37	3	3	1.79	86.05	6.98	6.98
总计	421	541	397	43	101	1.29	73.38	7.95	18.67

综合表 3、4、5 可知：《中图学报》、《大图学报》和《情报学报》所抽取样本刊载的论文总数分别为 409、460 和 421，比较接近；而数据引用平均值分别为 2.14、2.17 和 1.29，《中图学报》和《大图学报》的数据引用现象较《情报学报》要多一些；尽管三种刊物规范引用的比率在 2008 年前后均有了较大提升，但是非规范引用的现象仍然大量存在，只是《情报学报》的情况相对更好一些。

2.3 论文比较分析

2003—2013 年间图书情报学领域三种期刊论文数据引用的频次分布如图 2 所示。从图中可以看出，有数据引用行为和无数据引用行为的论文基本各占一半，可见数据引用作为一种常见引用行为，其普遍程度还有待提高。在有引用行为的论文中，引用次数在 5 以下的占绝大多数，说明目前图书情报学领域论文的数据引用频次相对较小。

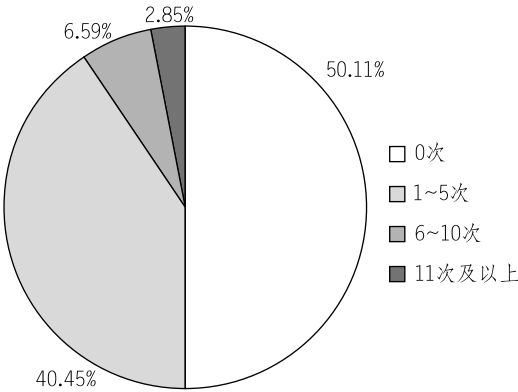


图 2 图书情报学领域论文数据引用频次分布图

表 6 中列出了 2003—2013 年间数据引用次数在 15 次以上的 18 篇论文，其中《中图学报》7 篇、《大图学报》9 篇，各约占半壁江山，《情报学报》仅有 2 篇。对 18 篇高数据引用论文进一步分析发现，技术特征较强的论文仅有一篇——《自举在词义消歧中的应用及其关键问题》，绝大多数高数据引用论文以资料提供为论文重要

表 6 高数据引用论文列表(数据引用次数≥15)

序号	论文	期刊	年份	数据引用次数
1	大学图书馆建筑的六十年变迁	大图学报	2010	41
2	电子书阅读器在国外图书馆的应用现状及存在问题	大图学报	2011	36
3	近年高校图书馆文献资源利用分析	大图学报	2003	26
4	公共图书馆“用地”与“建设”主要指标解析	中图学报	2009	25
5	自举在词义消歧中的应用及其关键问题	情报学报	2010	24
6	公共图书馆工作人员保障机制研究	中图学报	2010	24
7	国外网络引文研究的现状及展望	中图学报	2010	24
8	图书馆员去职业化问题、原因及对策研究	中图学报	2011	23
9	中美高职高专图书馆的比较分析	中图学报	2007	21
10	面向 Y 一代的大学图书馆用户服务	大图学报	2011	21
11	王肖珠与岭南大学图书馆	大图学报	2012	21
12	美国公共图书馆未成年人服务的特点	中图学报	2013	19
13	柳诒徵与江苏省立国学图书馆	大图学报	2012	18
14	俄罗斯文献呈缴制度研究	中图学报	2007	17
15	抄袭剽窃论文识别研究综述	情报学报	2007	16
16	近代中国教会大学图书馆文献收藏特点分析(下)	大图学报	2007	15
17	低利用率文献储存库的经济性研究	大图学报	2010	15
18	国外图书馆自动化系统市场发展状况研究	大图学报	2012	15

内容,而以数据为基础的文献计量学论文则未见踪影。究其原因,虽然文献计量学论文中存在大量数据,但这些论文往往都倾向于以丰富完备的方式提供和呈现个人通过大量数据得来的结果,而较少论及与公开数据的比较或对他人数据成果的借鉴。相应地,文献计量学方面的论文虽然提供了大量的数据结果,但是这些数据成果目前尚未被广泛引用。在某种程度上,这些论文在方法上似乎比数据成果吸引了更多眼球,人们对新颖或完美的研究方法和可视化效果的关注胜过对研究结果的期待。此外,数据被引最多的文献中有很多属于图书馆史的范畴,而引用这些数据的论文也因其丰富的资料内容展示而带有浓厚的史料色彩。

3 现状分析及对策

3.1 数据引用数量相对较少及原因

随着信息技术的进步和图书馆服务的拓展,学者具有更多的信息获取途径和更为便捷的信息获取方式,再加上优秀期刊对学术规范的要求日益严格,近年来,图书情报领域期刊的参考文献普遍较多。以2013年第4期《中图学报》为例,该期共刊登论文13篇,引文405条,平均每篇论文有31.15处引文,但其篇均数据引用仅为3.69次,大大低于总引文量。而表2的数据也显示出,11年间三大期刊的篇均数据引用仅为1.87次。

数据的价值毋庸置疑,而现实中骨感的数据引用数量至少反映了两方面问题。

(1) 大量的数据没有得到有效应用

信息社会产生了丰富数据,已经将我们的生活催生到大数据时代,但是面对海量数据,却更加令人产生无力感;信息丰富,情报匮乏;资源众多,结构混乱;信息繁杂,泥沙俱下。因此,尽管坐拥海量数据,学者们在从事科研活动时,仍然难以获得可靠的、系统的、完整的科研数据。在学术界,每年都有大批学者出于科研需要搜集大量数据并进行数据处理,将处理结果

和分析结果作为科研成果以论文形式公布。但是在我国少有学者公布科研成果的原始数据和处理过程,由于缺乏成熟的机构仓储和有效的激励措施,这些耗费大量人力物力、凝聚智慧的数据不能进入科研交流领域得到共享,而是作为学者的个人独有资产秘而不宣。

(2) 已有数据的可获得性较差

在我国,图书馆提供的文献资源依旧基本局限于纸质文献和数据库两种形式,文献类型依旧是十大情报源,比起传统的纸质文献,数据库对文献的揭示可谓更进一步,但是其对文献的描述也主要囿于题名、卷期、作者、机构等表面特征,对文献内容的揭示仍主要停留在关键词和引文的层面上,对文献数据的引用缺乏规范,统一的标引。结果是:一则不知哪种文献有自己需要的数据,查全率不高;再则只知数据结果,不知原始数据,无法进行比较研究或者后续研究。

机构知识库是伴随开放存取运动的发展而兴起的一种学术交流与资源共享的新模式^[8],是对各项知识成果进行记录、组织、集成和展现的知识仓储。理想的机构知识库不但应具有丰富的资源数量,更应具备丰富的文献类型,不但应包括公开出版的白色文献,还应包括未公开出版的灰色文献,如学者在科研过程中获取的原始数据和产生的数据成果。机构知识库本应成为数据上传、存储的工具和数据查找、引用的重要来源地,然而遗憾的是,目前国内大学机构知识库的内容主要以期刊论文和学位论文为主,其他类型的文献乃至数据所占比例极低。此外,很多机构知识库设置了严格的访问权限,不对外开放或者只部分开放其资源,以规避侵犯版权的风险,但是过度保护也会影响公开获取,制约机构知识库的学术交流和资源共享。

3.2 不规范数据引用大量存在及对策

由表2可知,图书情报领域数据引用中的非规范引用(包括部分引用和无引用)仍然大量存在,占到了数据引用总量的约1/3,其中部分引

用占 8.9%,无引用约占 23.8%。

一般而言,部分引用表明作者已经意识到了文献写作过程中出现了数据引用行为,并进行了来源说明,但由于对数据引用的规范性认识不足,再加上目前缺乏统一的数据引用格式,导致特征信息著录不齐全,标引来源时具有很大的随意性,以致无法通过作者的提示追溯到原始数据。一些致力于数据引用研究的机构和组织,如 DataCite、ANDS、DCC、Thomson Reuters 等,已经在一些问题上取得了重要进展^[9],提出了多种数据引用格式,如 Thomson Reuters 建议采取以下数据引用格式“作者(出版年份):数据标题.出版机构.访问地址”。例如:Crous, Pedro et al(2013): Phylogenetic lineages in Pseudocercospora. TreeBASE. <http://treebase.org/treebase-web/search/study/summary.html?id=12805>。但该数据引用格式能否得到一致认可尚待实践检验。尽管在具体的引用要素上仍然存在争议,但各机构已在数据引用的完整性、稳定性等原则上达成一致。

在不规范引用中,大部分情形下作者未给出任何标注信息,即无引用。有时是作者未意识到对数据的引用需要像文献引用那样进行著录,目前绝大部分学术期刊都有详细的文献引用方面的声明及格式要求,但专门针对数据引用的要求则较少。数据引用规范性的提高尚需作者、期刊和机构的共同重视和努力。有时则

是论文整体建构于某些统计数据的基础上,因此可能在论文的开头提及数据来源,但在正文中涉及数据细节时则无任何标注,使数据来源处于“默认式”或者“随你猜”的境地。针对这种情况,笔者建议介绍数据来源的时候加以明确声明,如“本文中所引用数据如无特定声明,均来源于……”。

数据引用著录格式的成熟和统一,以及数据引用规范共识的达成和推广,也将为数据的组织和检索提供支持和依据,从而促进数据引用行为的增长。

4 小结

综上所述,我国图书情报领域期刊的数据引用规范化程度总体来看呈逐步走高态势,但非规范引用仍然大量存在,且数据引用频次较少。目前制约数据引用行为、影响数据引用规范性的因素包括数据公开程度较低,已有数据的可获得性较差,作者对数据引用不够重视,缺乏统一的数据引用规范等。要推动数据引用行为的发展,就要建立相对稳定、规范、可靠的数据仓储;而要减少数据的不规范引用,就要制定科学规范的数据引用格式。同时,在政策上对数据共享予以肯定和奖励,也是推动数据引用向前发展的重要动力。

参考文献

- [1] Thomson Reuters. Collaborativescience: solving the issues of discovery, attribution and measurement in data sharing [EB/OL]. [2014-06-07]. http://wokinfo.com/products_tools/multidisciplinary/dci/collaborative_science.
- [2] 司莉,邢文明.国外科学数据管理与共享政策调查及对我国的启示[J].情报资料工作,2013(1):61-66. (Si Li, Xing Wenming. Scientific data management and sharing policies in foreign countries: investigation and inspiration to US[J]. Information and Documentation Services, 2013(1):61-66.)
- [3] 丁楠,潘有能.数据引用索引工作机理与应用现状综析[J].情报理论与实践,2014,37(6):59-62. (Ding Nan, Pan Youneng. Analysis of the working mechanism and application status of data citation index [J]. Information studies: Theory & Application, 2014, 37(6):59-62.)
- [4] White H D. Citation analysis of data files use[J]. Library Trends, 1982, 31(3): 467-477.
- [5] Valerie E, Sarah W, Nicholas M, et al. Data citation in the wild [EB/OL]. [2014-06-07]. <http://>

precedings. nature. com/documents/5452/version/1.

- [6] Hailey M, Mark P. The anatomy of a data citation: discovery, reuse, and credit[J/OL]. Journal of Librarianship and Scholarly Communication, 2012, 1(1): eP1035[2014-06-07]. <http://jlscl-pub.org/jlsc/vol1/iss1/6/>.
- [7] 墨愚. 数据引用的学术规范[J]. 编辑学刊, 2004(3): 68-69. (Mo Yu. Academic normalization of data citation [J]. Editors Bimonthly, 2004(3): 68-69.)
- [8] 万文娟, 吴高. 我国机构知识库内容建设问题与策略分析[J]. 图书馆, 2013(1): 110-114. (Wan Wenjuan, Wu Gao. Analysis on issues and strategies of institutional repository content construction in our country [J]. Library, 2013(1): 110-114.)
- [9] 侯经川, 方静怡. 数据引证研究: 进展与展望[J]. 中国图书馆学报, 2013, 39(1): 112-118. (Hou Jingchuan, Fang Jingyi. Review on data citation in the context of big data [J]. Journal of Library Science in China, 2013, 39(1): 112-118.)

丁楠 浙江大学图书馆参考咨询部馆员。

通讯地址: 浙江大学玉泉校区图书馆 513 室。邮编: 310027。

丁莹 浙江大学公共管理学院信息资源管理系本科生。

通讯地址: 浙江大学紫金港校区蒙民伟楼公共管理学院信息资源管理系。邮编: 310058。

杨柳 浙江大学公共管理学院信息资源管理系本科生。通讯地址同上。

凌晨 浙江大学公共管理学院信息资源管理系本科生。通讯地址同上。

潘有能 浙江大学公共管理学院信息资源管理系副教授, 硕士生导师。通讯地址同上。

(收稿日期: 2014-06-14; 修回日期: 2014-07-22)