

# 社会科学领域科学数据的引用现状与特点分析\*

屈亚杰<sup>1</sup>, 王亚男<sup>2</sup>

(1. 北京师范大学政府管理学院, 北京 100875; 2. 中央财经大学马克思主义学院, 北京 100081)

**摘要:** 从被引科学数据内容的视角分析社会科学领域科学数据的引用现状与特点。通过抽样调查法获取研究样本, 统计样本被引社会科学数据在创建者、类型、被引次数、访问方式、更新次数、规模及时间跨度等方面的概况。结果发现, 被引社会科学数据的创建者多是政府机构和研究机构; 被引社会科学数据的类型虽多样化, 但调查类数据居于主导, 被引用次数最多; 公众访问和仅ICPSR用户访问的样本数据集的数量相当; 大部分被引社会科学数据的时间跨度短、规模小、更新次数少。

**关键词:** 社会科学; 科学数据; 引用

**中图分类号:** G255

**DOI:** 10.3772/j.issn.1673-2286.2017.06.004

## 1 引言

引用是良好科学实践的基本要求, 一直是科学界及图书情报领域关注的重点问题。参考文献已经成为科技文献的重要组成部分, 科技文献发表或出版后被其他作者学习、参考、借鉴和引用, 辅助新文献的发表, 促进学科体系循序渐进地发展。随着计算机在科研方面大量应用, 科研过程产生海量科学数据。同时, 数据密集型科学研究范式的兴起使科学数据成为驱动科学研究发展的重要力量<sup>[1]</sup>。而发表的学术论文与支撑数据的断层会产生很多问题, 如科学研究成果的不可重现性、科学数据不可再利用、数据重复收集、相同数据子集的激增及科学欺诈等。因此, 科学数据的引用问题日益受到关注。众多学术机构提出, 科学数据和学术论文同样重要, 科学数据需要被正确引用。科学数据引用指作者在论文中以参考文献、脚注或文中注等方式, 对其所使用的科学数据标注来源出处<sup>[2]</sup>。规范科学数据引用是对数据集创建或管理者贡献的认可, 能够实现对基础数据的溯源, 便于验证科学研究推进; 可跟踪和评价数据集影响, 有助于科学数据的共享与传播。

社会科学领域数据涉及范围广泛, 多与社会发展、

民生国计密切相关。20世纪80年代, White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数据<sup>[3]</sup>。国外许多致力于共享科学数据的组织和机构广泛关注 and 推动科学数据引用<sup>[4]</sup>的发展。如2015年8月—2016年12月, 国际科技数据委员会数据引用与实践任务组在俄罗斯、澳大利亚、芬兰、以色列、印度、中国、日本等召开国际数据引用系列研讨会, 致力于促进数据引用联合声明和数据引用政策在全球范围内的应用<sup>[4]</sup>。鉴于统计工作量和数据库内容全面性等因素, 本文以美国校际社会科学数据共享联盟存储库为例, 选取ICPSR Bibliography of Data-Related Literature数据施引文献库的208条社会科学数据为研究对象, 从创建者、类型、引用次数、访问方式、更新次数、规模及时间跨度等维度展开分析, 以了解社会科学领域数据被引现状与特点, 为数据引用实践的进一步发展提供借鉴。

## 2 社会科学领域数据引用研究进展

通过文献调研发现, 目前直接研究社会科学领域数据引用的论文较少, 且均采用实证研究的方式; 而更多研究是从宽泛的科学数据引用角度展开, 尤其表现

\* 本研究得到国家社会科学基金项目“云计算环境下图书馆信息资源安全政策法律研究”（编号：11CTQ004）资助。

在理论研究方面。

理论方面的研究呈现现状综述与问题探讨相结合的特点。侯经川等归纳国际众多致力于数据引证相关研究成果,对国际研究现状进行梳理,并提出亟待解决的三个问题:数据引证现状评估与研究进展追踪,数据引证索引编纂、指标体系研究及数据引证数据库建立,基于数据引证行为、记录及索引分析<sup>[5]</sup>。屈宝强等从科学数据引用实践、各利益相关者态度、科学数据引用格式、基于引用的数据发现与评价等维度对数据引用现状进行分析,并指出科学数据引用虽已引起学术界重视,但仍存在缺乏对数据引用的氛围和意识、没有通用的数据引用格式和位置指南等问题<sup>[6]</sup>。而张静蓓等在总结数据引用所面临关键问题的基础上,从数据引用规范发展、文献管理软件调研、数据引用利益相关的责任与义务三方面概括目前国外数据引用的研究进展,并指出数据共享和引用意识缺乏、引用对象不明确及数据如何长期保存是数据引用面临的三大难题<sup>[7]</sup>。

对于社会科学领域数据引用的论文一般选择期刊作为样本,涉及学科包括图书情报学、社会学、人口学等。Sieber等对心理学、社会学及人类学等学科的1 000篇论文内容进行分析后,发现很多作者没有对数据引用的原始数据作者和数据来源出处进行说明,只有19%的作者在参考文献中提到引用的数据<sup>[8]</sup>。丁楠等分别对我国图书情报领域(2003—2013年)和社会学领域(2003—2014年)代表性期刊的数据引用行为进行调查分析,发现规范引用比例虽总体呈现逐渐升高的状态,但非规范引用比例仍不容小觑<sup>[9-10]</sup>。Mooney等通过Wilson Web数据库检索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,并按人文科学、社会科学和自然科学对检索结果随机抽样,选取分属于44种期刊的65篇论文作为研究样本<sup>[11]</sup>。研究发现,约有一半期刊刊发过数据引用方面的声明及格式要求,但论文中很少有对其引用数据进行注释。

以上研究主要关注引用行为本身,即论文对所使用科学数据的标引状况,而较少涉及被引科学数据内容。因而本文从被引科学数据内容的视角展开研究,通过分析其不同维度的统计数据,总结社会科学领域科学数据的引用现状与特点。

### 3 数据来源与研究方法

美国校际社会科学数据共享联盟存储库(Inter-

university Consortium for Political and Social Research, ICPSR)由美国密歇根高校社会研究中心于1962年建立,是国际重要的学术研究数据资源库之一。其创立目的是收集与保存社会科学研究数据资源,提供公开并公平的使用渠道,以提升数据资源的利用率<sup>[12]</sup>;主要任务是维护并提供研究、教学所需的社会科学数据资料档案并提供训练与协助,使研究者在量化研究中能有效运用资料。ICPSR现维护50余万条社会科学和行为科学研究数据档案,9 000多个数据集,包括教育、老龄化、刑事司法、药物滥用、恐怖主义等21个专业数据集,是目前世界上最大的社会科学数据中心。而ICPSR Bibliography of Data-Related Literature是一个可搜索的数据施引文献库,收集与ICPSR存储和发布的数据资料集相关的引用文献资料,截至2016年该数据库已收录超过7万篇引用科学数据的论文,涉及6 752条科学数据。

本文以ICPSR Bibliography of Data-Related Literature数据库为例,对其收录论文所引用的社会科学数据进行抽样研究,分析样本被引社会科学数据在创建者、类型、被引次数、访问方式、更新次数、规模及时间跨度等方面的分布情况。首先,将ICPSR Bibliography of Data-Related Literature数据库收录的6 752条数据集按照首字母A—Z进行排序,按结果显示的顺序每30条数据集归为一组,共分成225个组;其次,采取等距抽样方法,每组选取第30条数据集作为样本,获得初步样本225个;最后,依据元数据项的完备程度(包括数据集的创建者、类型、引用次数、访问方式、更新次数、规模及时间跨度)对样本进一步筛选,获得样本数据集208个。由于ICPSR存储库是世界最大的社会科学数据中心,样本数据采用科学的抽样方法获得,因而具有一定代表性。

## 4 社会科学领域科学数据的引用现状分析

以ICPSR为例,从科学数据内容的视角分析社会科学数据的引用现状。分析维度包括被引社会科学数据的创建者、类型、引用次数、访问方式、更新次数、规模及时间跨度等。

### 4.1 被引社会科学数据创建者

社会科学数据创建者,指对社会科学研究作出主

要贡献,并能授权数据存储库保存和发布研究所产出科学数据的科研人员或机构。因而,从宏观层面看,社会科学数据的创建者分为个人创建者和机构创建者。其中,个人创建者主要是科研人员,而机构创建者包括政府机构、研究机构(高校和研究所)、企业、协会和基金会等。被引社会科学数据的个人创建者和机构创建者数量统计情况如表1所示。

表 1 被引社会科学数据的不同创建者数量统计

	个人创建者	机构创建者			
	个人	政府机构	研究机构	企业	协会和基金会
被引社会科学数据集量/个	33	98	56	17	4

通过对208个样本数据集的创建者分析,发现33个被引社会科学数据集来源于个人创建者(占比16%),有175个被引社会科学数据集来源于机构创建者(占比84%)。机构创建者可分为政府机构、研究机构、企业、协会和基金会四类,其中政府机构创建者有98个(占比56%),是被引用科学数据的主要产出机构,包括美国人口普查局、美国司法统计局、联邦司法中心、美国国家教育统计中心、美国劳工统计局、美国国家卫生统计中心、美国社会保障局、欧盟委员会等。政府机构所创建的数据多与经济发展、民生需求密切相关,开放程度较高,因而很容易被科研人员引用。例如, *American Housing Survey, 1995: MSA Core and Supplement File* 涉及9个大都市统计区域的住房信息(包括建造年份、住房类型和人数、商业或医疗机构设施的完善度等),该数据的创建者是美国人口普查局<sup>[13]</sup>。研究机构包括高校和研究所,该类机构创建者有56个数据集(占比32%),是科学数据的重要创建者和拥有者,包括密歇根大学、威斯康星大学、约翰霍普金斯大学、加州大学伯克利分校、弗吉尼亚联邦大学、埃塞克斯大学、美国城市研究所、维拉司法研究所等。例如, *Health Reform Monitoring Survey, First Quarter 2015* 是美国2015年第一季度卫生改革监测的社会科学数据,调查内容包括健康状况自我报告、获得医疗保险和支付医疗费用的困难、健康保险类型以及没有健康保险的原因等,数据创建者是美国城市研究所<sup>[14]</sup>。企业、协会和基金会虽然也是科学数据的创建者,但由于企业的营利性、协会和基金会很少直接开展研究等因素的限制,这两类机构创建者所产出的科学数据

很少被其他研究者引用。

## 4.2 被引社会科学数据类型

对208个样本数据集的类型进行调查,发现ICPSR将被引的社会科学数据主要分为调查数据、行政记录数据、事务数据、综合性数据、普查数据和其他数据(观察数据、实验数据、临床数据、医学记录等)类型。其中,调查数据集数量最多(130个)、行政记录数据集27个、事务数据集22个、综合性数据集13个、普查数据和其他类型数据集各8个。各类型社会科学数据所占比例如图1所示。

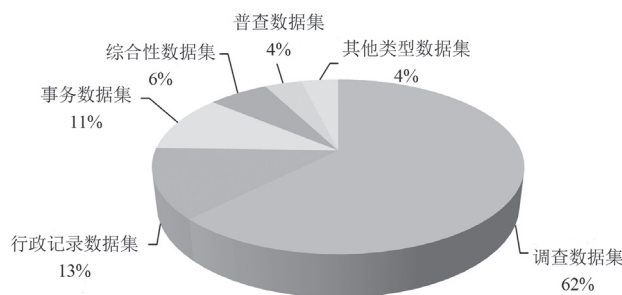


图 1 被引社会科学数据中各类型数据集占比情况

调查数据在样本数据集中占比62%,是重要的社会科学数据类型。与自然类学科通常需要依靠实验室、仪器、设备等产出科学数据的方式不同,社会科学一般采用调查法获取研究数据。例如,密歇根大学为衡量消费者态度和期望变化,了解发生变化的原因,对影响消费者信心的不同因素展开月度调查(包括个人财务、就业、物价变动、国家商业状况,以及种族、性别、年龄和教育等)<sup>[15]</sup>,该调查数据对预测总体消费者行为的变化至关重要。行政记录数据集和事务数据集占比分别为13%和11%。行政记录数据是国家行政部门为实现服务、监督和管理目的,通过对系统内的登记对象收集相关信息和对信息进行加工所形成的行政记录统计化后的数据<sup>[16]</sup>。行政记录数据是政府统计数据的重要来源,能有效解决政府统计数据获取困难的问题。事务数据是描述具体事件的数据,一般具有时间维度和数值维度,可指代一个或多个事件对象(如存储记录、活动记录等)。这两类数据的创建者多为政府部门,随着政务公开实践的发展,这些数据的共享程度逐步提高,因而相比其他类型数据,行政记录数据和事务数

据被引用的数量仅次于调查数据。综合性数据、普查数据和其他类型数据（观察数据、实验数据等）的总占比为14%。其中，综合性数据指通过一定方法将单个数据进行集成后形成的数据，在某种意义上，综合性数据的表现形式是汇总表格。如美国人口普查局的“摘要文件”属于综合性数据，包括在特定地理区域具有各种指定属性的个体总数<sup>[17]</sup>。普查数据指国家或地区为详细调查某项重要国情，专门组织的一次大规模全面调查而产生的原始数据，普查数据通常被用来说明现象在一定时点上的全面情况，如人口普查数据、农业普查数据、经济普查数据。

### 4.3 社会科学数据被引次数

分别从时间和数据类型的维度对208个样本数据集的被引用情况进行统计，如表2和图2所示。2011—2015年调查数据的被引总数是342次，明显高于其余数据类型的被引总数，说明科研人员对调查数据的需求较大。另外，调查数据的总被引次数较高与ICPSR存储库收录和发布的调查数据数量较多密切相关，便于科研人员及时获取和使用。而2011—2015年普查数据的被引次数为3次，是所有社会科学数据类型中引用次数最少的，可能因普查数据通常需要长周期的调研才能获取，且其规模一般很大，通过存储库发布所需时间较长，使得ICPSR存储库收录和发布的普查数据数量较少。2011—2015年社会科学数据的总被引次数均超过95次，并在2013年达到峰值，2014年和2015年有所下降，但依然比2012年的被引次数多。这说明，ICPSR存储库所收录和发布的社会科学数据对科研人员的影响力逐渐增强，而科研人员对社会科学数据的引用意识也在不断提高。

### 4.4 被引社会科学数据访问方式

由于ICPSR所发布的社会科学数据访问权限包括面向公众和仅面向ICPSR用户两种方式，因此，被引社会科学数据的访问方式分为直接访问和登陆ICPSR才能访问。对208个样本数据集的访问方式进行统计，被引社会科学数据样本中直接访问和登陆ICPSR访问的数量分别为103个和105个，说明ICPSR存储库所发布的社会科学数据共享程度较高，科研人员访问时不需要与ICPSR成员机构联系。对于登陆ICPSR才能访问的社

表 2 2011—2015年数据集不同分类型被引情况

数据集类型	2011年	2012年	2013年	2014年	2015年
调查数据	51	58	63	76	94
行政记录数据	11	29	64	14	10
事务数据	5	6	10	17	14
综合性数据	17	15	14	19	15
普查数据	0	1	1	0	1
其他数据	14	7	10	11	3

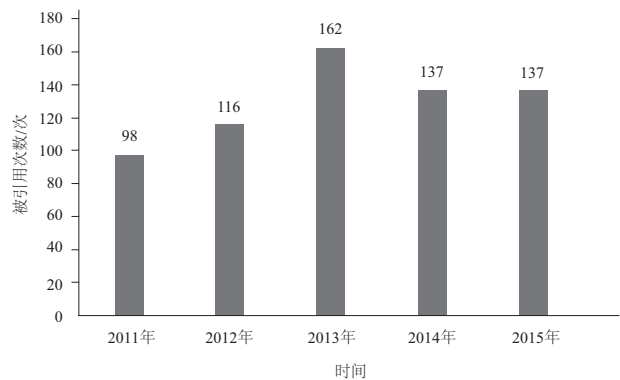


图 2 2011—2015年数据集被引总数情况

会科学数据，只要科研人员所在机构是ICPSR的成员机构或获取到ICPSR成员机构内科研人员的账号，均可访问。

### 4.5 被引社会科学数据更新次数

更新次数指自社会科学数据首次发布后的修改次数。更新次数越多，社会科学数据的版本越多，说明内容变化越大。对208个样本数据集的更新次数进行统计，发现没有更新的被引社会科学数据量是99个，占比47%，说明样本中有将近一半的被引社会科学数据只有一个版本；目前有2个及以上版本的被引社会科学数据在样本中占比53%，更新1次的占比35%，更新2次的占比10%，更新2次以上的占比8%。被引社会科学数据更新次数最多达到8次，调查样本中仅有2个。总体来看，对数据的更新包括新增、更改、删除等操作。如美国卫生与人类服务部通过《国家药物滥用治疗服务调查（2000年）》产出的调查数据目前有9个版本，更新内容包括添加变量、删除变量、调整变量顺序、添加相应的Stata数据定义语句、更新XML文件、修订PDF文档等<sup>[18]</sup>。

## 4.6 被引社会科学数据规模

被引社会科学数据的规模指其所占内存大小。数据规模可能与数据创建主体、数据类型、数据时间跨度等因素有关。对208个样本数据集的规模进行统计,发现规模小于50MB的样本数据集有124个,占比59%,说明样本中超过一半的被引社会科学数据的规模较小。被引社会科学数据规模大于50MB的占比41%,其中50—500MB的被引社会科学数据量有66个,占比32%;数据规模大于500MB的有18个,占比低于10%。在18个数据规模大于500MB的被引社会科学数据集中,有17个由政府机构创建,说明政府机构所产出并发布的社会科学数据规模通常较大。其中,规模最大的被引社会科学数据是美国商务部和人口普查局于2004年产出的调查数据,该数据包含18个数据集,数据规模达到51670MB,详细反映了美国家庭和个人经济状况<sup>[19]</sup>。

## 4.7 被引社会科学数据时间跨度

被引社会科学数据的时间跨度,指其所揭示内容所包含的年限。对208个样本数据集的时间跨度进行统计,发现样本中时间跨度在1年内的被引社会科学数据量是124个,占比60%,说明大部分被引社会科学数据揭示的是具体某年的信息。如密歇根大学社会研究所开展的研究揭示出2002年美国青少年关于重要价值观、行为和生活方式方面的内容<sup>[20]</sup>。样本中被引用社会科学数据的时间跨度大于1年的数量是84个(占比40%),其中时间跨度1—5年的被引社会科学数据占比29%,大于5年的被引社会科学数据占比为11%。时间跨度最大的被引社会科学数据由美国高校政治和社会研究联盟创建,该数据时间跨度长达181年,包含关于美国社会和政治方面的大量数据,并按州、种族、国籍、家庭人数、家庭规模、出生、死亡、婚姻、职业、宗教和一般经济状况等进行分类<sup>[21]</sup>。

## 5 社会科学领域被引科学数据的特点

通过分析社会科学领域科学数据的引用现状,发现其具有数据创建者多是政府机构和研究机构,数据类型多样、调查类数据被引次数较多,公众访问和仅ICPSR用户访问的样本数据集的数量相当,大部分被引社会科学数据的时间跨度短、规模较小、更新次数少等

特点。

(1) 被引社会科学数据的创建者多为机构创建者。与被引科学数据的个人创建者相比,机构创建者明显居于主导地位。同时,被引社会科学数据的创建者多是政府机构和研究机构。机构创建者产出和发布的研究数据被引用量较多的原因可能有两方面:一方面,在科研人员开展研究时申请国家基金资助不容易,受时间、资金及精力等方面限制,研究成果产出周期较长、共享程度有所欠缺;另一方面,机构创建者开展的研究多依托国家基金资助项目,研究内容范围广泛,研究团队协同工作,产出的研究成果可信度和共享程度较高。政府机构在机构创建者中占比56%,其创建的社会科学数据被引用程度较高。这是因为政府数据与民生国计密切相关,是社会科学领域开展研究所需数据的重要来源,同时美国政府数据开放程度较高也是重要因素。另外,高校和研究机构开展研究时通常会有科研资助机构支持,而科研资助机构则要求其资助产出的科研成果及科学数据能及时以适当的方式共享,促进科研成果效益最大化(如英国研究理事会最主要的任务是资助研究)。《RCUK数据政策通用原则》规定,由其资助项目所产生的科学数据是公共物品,为公众利益而产生,应在不损害知识产权和隐私的情况下,尽可能及时开放共享<sup>[22]</sup>。

(2) 被引社会科学数据的类型多样,调查类社会科学数据的被引次数较多。由于创建主体和研究方法的不同,被引社会科学数据包括调查数据、行政记录数据、事务数据、综合性数据、普查数据等多种类型,其中调查类数据在总数据集中占比最大,这说明对于社会科学研究而言,调查数据起着举足轻重的作用。例如,对于同类研究的后续研究者而言,引用关于该主题的已有调查数据,以避免重复劳动,节省研究成本。同时,拟引用调查数据的确定性和可参考性也尤其重要。确定性主要表现为调查数据指向明确的时间与对象,可参考性指调查数据具有代表性、调查数据与研究主题具有一致性。在通过电话访谈、个人访谈、问卷调查及网络调查等方法开展研究时,确保数据的确定性和可参考性十分重要。

(3) 公众访问和仅ICPSR用户访问的样本数据集的数量相当。引用科学数据的前提是科学数据可获取,因而,科学数据的访问方式对引用具有重要影响。由于受存储库的属性限制,不同存储库的公开访问程度不同。如国家数据中心一般面向公众访问,而机构存储库部

分提供公众访问。因此,一方面数据存储库应尽可能提供公众访问途径,以促进科学数据再利用;另一方面,科研人员可通过多种途径访问和获取研究所需数据。

(4) 大部分被引社会科学数据的时间跨度短、规模较小、更新次数少。社会科学数据的时间跨度短意味着其产出的周期短、规模小、效率高,使得数据上传、下载比较方便;而更新次数少则表明社会科学数据的版本少,内容变化小。因此,该类社会科学数据更易被研究者引用。

## 6 结语

大数据时代数据的价值日益受到科研人员的重视,数据共享实践的发展也使得社会科学数据更容易被科研人员获取并引用。本文基于对ICPSR Bibliography of Data-Related Literature数据库被引社会科学数据概况的统计,从被引科学数据的创建者、类型、引用次数、访问方式、更新次数、规模及时间跨度等7个维度分析引用现状,并总结被引社会科学数据的4个特点,有助于科研人员明晰目前社会科学领域数据的被引情况,为数据引用实践的进一步发展提供借鉴。

研究尚存在一些不足,例如样本量不大,本文采用等距抽样方法,仅抽取208个样本数据集展开统计分析。另外,本文仅选取社会科学领域被引数据的现状与特点展开分析,未涉及自然科学领域的科学数据,也没有深入到具体某一学科,样本来源数据库属于美国。因此,下一步研究方向可侧重于自然科学领域科学数据或某一具体学科数据,也可以选择其他国家(如英国、澳大利亚)的数据存储库作为研究对象,开展相关研究。

## 参考文献

- [1] 吴金红,陈勇跃.面向科研第四范式的科学数据监管体系研究[J].图书情报工作,2015,59(16):11-17.
- [2] 丁楠,黎娇,李文雨泽,等.基于引用的科学数据评价研究[J].图书与情报,2014(5):95-99.
- [3] WHITE H D. Citation analysis of data files use[J].Library Trends, 1982,31(3):467-477.
- [4] CODATA.International Series of Data Citation Workshops[EB/OL].[2017-04-01].<http://www.codata.org/task-groups/data-citation-standards-and-practices/international-series-of-data-citation-workshops>.
- [5] 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(1):112-118.
- [6] 屈宝强,王凯.科学数据引用现状和研究进展[J].情报理论与实践,2016(5):134-138,114.
- [7] 张静蓓,吕俊生,田野.国外科学数据引用研究进展[J].图书情报工作,2014(8):91-95,138.
- [8] SIEBER J E,TRUMBO B E.(Not) giving credit where credit is due:citation of data sets[J].Science and Engineering Ethics,1995,1(1):11-20.
- [9] 丁楠,丁莹,杨柳,等.我国图书情报领域数据引用行为分析[J].中国图书馆学报,2014(6):105-114.
- [10] 丁楠,杨柳,丁莹,等.我国社会学期刊论文数据引用行为研究[J].图书与情报,2014(6):88-93.
- [11] MOONEY H,NEWTON M P.The anatomy of a data citation: discovery,reuse,and credit[J].Journal of Librarianship and Scholarly Communication,2012,1(1):1035.
- [12] ICPSR.About ICPSR[EB/OL].[2017-04-01].<https://www.icpsr.umich.edu/icpsrweb/content/about/>.
- [13] ICPSR.American housing survey,1995:MSA core and supplement file(ICPSR 2368)[EB/OL].[2017-05-08].<http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/2368>.
- [14] ICPSR.Health reform monitoring survey,first quarter 2015(ICPSR 36364)[EB/OL].[2017-05-08].<http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/36364>.
- [15] ICPSR.Survey of consumer attitudes and behavior,March 1996(ICPSR 2943)[EB/OL].[2017-05-08].<http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/2943>.
- [16] 方越峦,刘建平.行政记录:政府统计不可忽略的数据资源[J].统计研究,2013(10):22-29.
- [17] ICPSR.Glossary of social science terms[EB/OL].[2017-04-01].<http://www.icpsr.umich.edu/icpsrweb/ICPSR/support/glossary>.
- [18] ICPSR.National survey of substance abuse treatment services (N-SSATS),2000(ICPSR3436)[EB/OL].[2017-04-01].<http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/3436>.
- [19] ICPSR.Survey of income and program participation(SIPP)2004 panel(ICPSR 4517)[EB/OL].[2017-04-01].<http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/4517>.
- [20] ICPSR.Monitoring the future:a continuing study of American youth(8th- and 10th-grade surveys),2002(ICPSR 3752)[EB/OL].[2017-04-01].<http://www.icpsr.umich.edu/icpsrweb/ICPSR/>

studies/3752.

www.icpsr.umich.edu/icpsrweb/ICPSR/studies/3.

[21] ICPSR. Historical, demographic, economic, and social data: the United States, 1790-1970 (ICPSR 3) [EB/OL]. [2017-04-01]. http://

[22] Research Councils UK. RCUK common principles on data policy [EB/OL]. [2017-04-01]. http://www.rcuk.ac.uk/research/datapolicy/.

## 作者简介

屈亚杰, 女, 1991年生, 硕士研究生, 研究方向: 科学数据管理, E-mail: 819468068@qq.com。

王亚男, 女, 1989年生, 硕士, 研究方向: 信息分析。

## Research on the Citation Present Situation and Characteristics of Scientific Data in Social Science

QU YaJie<sup>1</sup>, WANG YaNan<sup>2</sup>

(1. The School of Government, Beijing Normal University, Beijing 100875, China;

2. The School of Marxism, Central University of Finance and Economics, Beijing 100081, China)

**Abstract:** Research on the citation present situation and characteristics of scientific data in social science from the perspective of the content of cited scientific data. The sample was obtained by sampling survey method, then statistics the overview of samples from the aspects such as creator, type, citation, the number of updates, size, time span and so on. After analyzing the statistical data, can draw the following conclusions. The creators of cited scientific data are mostly government agencies and research institutions. Although the types of cited scientific data are diversified, the survey data is dominant and the most cited. Public access and only ICPSR user access are equally divided. Most of the cited scientific data has short time span, small scale and less updated time.

**Keywords:** Social Science; Scientific Data; Cite

(收稿日期: 2017-04-13)

## ■ 书 讯 ■

# 《汉语主题词表》(工程技术卷)

《汉语主题词表》自1980年问世以后,经1991年进行自然科学版修订,在我国图书情报界发挥了应有的作用,曾经获得了国家科学技术进步二等奖。为了适应网络环境下知识组织与数据处理的需要,2009年由中国科学技术信息研究所主持,并联合全国图书情报界相关机构,完成《汉语主题词表(工程技术卷)》的重新编制工作。

全书共收录优选词19.6万条,非优选词16.4万条,等同率0.84。在体系结构、词汇术语、词间关系等方面进行改进创新。为了方便工程技术领域不同专业用户使用,《汉语主题词表》(工程技术卷)按专业分13个分册出版,同时建立《汉语主题词表》服务系统,提供在线概念检索和辅助标引服务,通过可视化技术展示各类概念关系,是图书馆、档案馆、出版社、期刊杂志社、文献信息中心等专业工作者及科研、教育及工程技术领域人员必备的参考书。

《汉语主题词表(工程技术卷)》已于2014年由科学技术文献出版社出版,全书2300余万字,总定价3880元,可分册购买。