

科学数据共享研究的文献计量分析

张翔

(1. 中国科学技术大学管理学院, 安徽合肥 230001)

摘要: 科学数据共享是促进现代科学发展重要的推动力量, 是推动国家科技创新的关键要素。以“科学数据共享”作为主题词, 不限定时间范围, 选择文献类型为学术期刊, 通过对《中国知网学术期刊网络出版总库》的文献检索, 检索到234篇相关的学术论文, 综合利用SATI3.2、Ucinet6与SPSS18.0等软件对样本文献的年度发文量、期刊分布、核心作者分布、研究机构分布、科研基金资助情况以及研究主题等8个方面进行分析, 从而为国内相关研究提供一定的参考。

关键词: 科学数据共享; 文献计量分析; 共词分析

中图分类号: G201

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2013.03.011

Statistical Analysis of Literatures for the Scientific Data Sharing Research

Zhang Xiang

(School of Management, University of Science and Technology of China, Hefei 230026)

Abstract: The scientific data sharing is an important driving force promoting the development of modern science, which is the key element of promoting national scientific and technological innovation. Subject headings by "Science Data Shared" for academic journals without time range, searches 234 articles correlative academic paper through on the CNKI Databases under of literature retrieved in this paper, and then it analysis annual issued volume, journals distribution, core author distribution, research body distribution, science research fund funding situation and eight theme area by the SATI3.2, Ucinet6 and SPSS18.0software on samples literature, so as to provide some reference to related studies in China.

Keywords: scientific data sharing, statistical analysis, Co-word analysis

科学数据共享在现代科学技术发展的过程中发挥着重要推动作用, 具有重大的经济价值、科学贡献以及社会效益^[1]。随着知识经济深入发展, 科学数据共享的重要性日益凸显, 通过彼此之间信息或数据的交换、共享与整合, 可以更好地推动科技创新的发展。国家科技创新的源泉来自于科学数据, 科技创新的前提是对现有科学数据的充分提炼与运用, 特别是新技术、新思想的产生必然依托于大量真实、可靠的科学数据才能进行真伪的鉴别与应用的扩展, 而科学数据也在共享及运用的过程中进行

补充与修正^[2]。当代科学技术发展的趋势已经呈现出大科学、交叉研究、量化以及注重研究过程的特点, 越来越依托于系统化的基础科学数据以及衍生出的其他数据产品, 而科学数据共享正是科学研究重要的资源提供平台^[3]。所以, 基于未来科学发展以及国家科技创新的需要, 科学数据共享研究是一项长期性、基础性、保障性的重要工作。目前, 国内学者从不同的视角对科学数据共享展开了研究, 已经取得了一定的研究成果。

作者简介: 张翔(1989-), 男, 中国科学技术大学管理学院硕士, 研究方向: 企业家精神与领导力。

基金项目: 2012年国家自然科学基金面上项目“评价理论与方法”(71222106)。

收稿日期: 2013年3月27日。

1 相关文献论题

国内学者基于情报学、计算机科学、管理学等不同的视角讨论了科学数据共享的原则、标准、数据库以及共享平台体系等相关内容,归纳如下。

(1)科学数据共享的原则与标准。李娟、刘德红与江洪以《网络时代的科学原则》与《公共资金资助的研究数据获取原则与指南》为实例,分析了共享原则制定的重要性及其在实际共享工程中的应用性,提出共享的公益性原则及开放性原则^[4]。王巧玲、钟永恒与江洪详细介绍了《布加勒斯特宣言》的主要内容,明确了确保获得信息与知识的权利、以可承受的费用水平实现信息普及化、促进语种多元化及文化主体意识等7项原则,充分体现了“以人中心”、“合作开放”的思路^[5]。黄鼎成、李晓波与莫纪宏基于《科技进步法》的要求、国际相关公约的约束,设计了我国科学数据共享的6项法律原则,包括合法性、共用性、公平竞争、产权保护、宏观调控、比例及责任原则^[6]。总之,科学数据的共享原则既包括经济原则、社会原则、法律原则以及技术原则等内容,是建立科学数据共享标准、体系、数据库、共享平台的根本性思路^[7]。国内学者针对共享的标准,从不同的学科背景或行业背景展开论述,例如朱星明、白婧怡与蔡佳男阐述了水利科学数据共享的标准^[8];刘德培与尹岭则构建了中国医药卫生科学数据的共享标准^[9];其他还有国土资源科学数据、地震科学数据、环境科学数据、工程科学数据等相关标准^[10-12]。共享标准的设计与说明主要侧重技术标准的内容,同时也涉及到法律标准、管理标准的建立及应用。

(2)科学数据共享的体系与元数据的应用。科学数据共享的体系包括了基本数据库的建设、完整系统的构建、服务平台的建立与管理等相关内容。聚焦最多的是有关元数据在科学数据共享系统中的应用与实现路径。李纪宾、邹小农与乔友林从元数据的内涵和标准分析了其在医学科学数据共享体系中的应用价值^[13]。王卷乐、诸云强与谢传节基于元数据能够整合分散的地学数据资源的基础上,设计了一套分布式的数据交换体系,构建了地学科学数据的共享网络平台^[14]。陈燕、吕晓春与孙静涛利用元数据的特性解决了焊接科学数据共享平台的关键问题——分散用户需求的整合问题^[15]。李雪山与熊燕基于元数据和WebService等计算机技术的应

用,设计了交通科学数据共享平台的方案,能够有效提高交通系统的服务能力^[16]。安基文、庄大方与袁文提出元数据及元数据标准是建立资源环境科学数据的共享系统的关键技术^[12]。由此可见,元数据在科学数据共享体系的建设过程中具有重要的应用价值。除此之外,也有部分学者从服务模式、框架设计、大数据处理能力、网络分布体系、网站界面平台等进行了分析,例如张小红与李思经从服务保障体系的视角探讨了共享网络、共享平台中的服务方式、服务策略、服务原则,提出农业科学数据共享平台的设计应该更多考虑农业生产者为对象的信息服务网络^[17]。陈腾与张子平通过对国内外相关资料的整理,提出了国土资源科学数据共享平台的层次、规模、分布结构等框架内容^[18]。黄鼎成、李晓波与王卷乐从国家视角提出未来5年时期内不同部门或行业中科学数据共享工程的战略规划与发展布局^[19]。总体而言,科学数据共享体系既涉及宏观层次的战略布局,也涉及到不同行业或部门内科学数据共享路径的实现。

(3)科学数据共享的机制与未来挑战。科学数据共享的机制是对科学数据共享过程中的各基本要素(共享技术、科学数据资源、共享规则、用户需求、组织管理、共享网络)的优化组合,从而实现科学数据的有效、及时、全面的共享,推动科学技术的积极进步。科学数据共享的机制主要涉及协同建设机制、持久运行机制、管理监督机制、基本保障机制等内容,涵盖了组织管理、系统运行、监督控制、条件保障等方面的具体操作^[20]。

朱雪忠与徐先东通过分析科学数据共享与知识产权之间的冲突,提倡政府在科学数据共享过程中,应该发挥主导作用,建立规范的产权交易机制和法律监督机制^[21]。路鹏等通过全国大范围的数据调查,发现国内在共享物质基础建设方面有较大进步,但是科学合理的运行机制、健全的组织管理制度尚未完全形成,需要在共享机制方面进行强化,实现科学数据共享的软硬条件的同时具备^[22]。朱星明等通过借鉴国外发达国家的经验,提出我国水利科学数据共享机制的构建模式,并认为政府、企业、科研机构之间的协同合作、产学研合作模式的推广,将会对科学数据共享事业的发展有重要的推动作用^[23]。孙晨晔与朱德米在网络治理理论的基础上,针对长三角地区先进制造业的科学数据共享现状,提出了外向式的网络治理共享模式^[24]。刘

闯提出科学数据的共享激励机制是科学数据队伍稳定、协作的关键^[25]。胡海燕与刘世洪则阐述了农业科学数据共享服务机制的内涵、特征与重要性^[26]。未来的时代是“大数据时代”，科学数据的信息量是巨大的，多学科、跨行业的数据交流趋势日益明显，数据形式日益多元化、数据处理的复杂度不断提高，这些要求了科学数据共享要有更加海量的信息存储、更加高效的信息处理、更加密集的信息交流，对多地区、多行业、多学科不同的科学数据进行有机整合，从而形成一个覆盖范围广、信息动态性强、主体参与度高的共享网络体系^[27-28]。

(4) 科学数据共享研究的方法类型。学者利用描述性研究、个案研究、数理实证、调查研究、文献分析等多种方法展开论证。例如，李思经与张小红通过对农业科学数据共享现状的分析，归纳、总结出有共享存在的问题，并提出若干建议^[29]；马海龙、陈佑启与邹金秋以“国家农业科学数据共享中心农业区划数据分中心”为实例，介绍了共享平台的系统设计与建设管理经验^[30]；李莉利用浙江疾控中心的大数据进行实证研究，总结出一种海量的数据管理模式^[31]；路鹏等通过调研的方式，梳理了我国科学数据共享的现状以及制约我国科学数据共享的主要因素^[22]。王广华以文献综述的形式全面总结了国土资源科学数据共享的研究进展与未来趋势^[10]。另外，也有部分学者利用文献计量分析、实验法等展开讨论，但是此类研究方法在该命题研究中并不多见。

2 文献统计分析说明

本文选取《中国知网学术期刊网络出版总库》作为文献来源，以“科学数据共享”作为主题词，并且以“数据共享”或“共享”作为题名词，不限定时间范围，选择中英文扩展检索，并确定文献来源的类型为期刊文献，一共检索到269篇期刊文献，并对初步检索的文献进行阅读，剔除会议报道、书评、研究简讯、学者报道等非学术性或者与主题关联度不高的文献，最后得到符合要求的234篇学术论文作为最终样本。笔者利用SATI3.2软件和SPSS18.0软件对样本文献进行处理，分析该领域研究的年度文献变化量、期刊分布、核心作者、高被引文献、研究机构分布、科研基金资助以及研究主题情况。

在统计样本文献的年度变化量时，分为全部

文献的年度变化与核心期刊论文的年度变化两个指标，核心期刊论文选取的标准为2012年北京图书馆的中文核心期刊、2011年CSCD收录期刊（包括扩展版）、2012年CSSCI收录期刊（包括扩展版）、EI或者SCI收录的国内期刊。关于作者分布的统计，除了反映核心作者情况外，同时通过SATI3.2软件来构建作者共现的二值矩阵，并利用Ucinet6软件输出的共现网络来表现国内学者之间相互合作的关联情况。高被引文献的分析是基于CNKI文献分析系统的数据报告，而研究主题的统计则是通过高频关键词的聚类分析而得出的结果，并以SPSS18.0输出的聚类树状图形式进行展示。

关于样本文献统计分析的结果见以下各节论述。

3 论文数量与期刊来源

3.1 论文数量

国内科学数据共享的研究主要兴起于21世纪^[32]，从样本文献的时间分布来看，2002年到2012年一共发表了234篇学术期刊论文，其中有93篇核心期刊的学术论文（CSSCI核心为22篇、CSCD核心为34篇），论文数量的年度变化呈现出波动的趋势，如图1所示。

由图1可以看出，自2002年到2008年，期刊学术论文总数呈现出总体增长的态势，而2009年以后有明显的下降趋势，而核心期刊论文数的变化规律也与之相同，但是增长或者减少的幅度比较小。其中，2008年发表论文数最多，发表论文数为39篇，占样本文献的17%，核心论文发表最多的则为2007年，一共有15篇，占样本文献的7%，为核心论文总数的17%。总体看来，2007年、2008年、2009年这3年为国内科学数据共享研究的高峰阶段。仅从论文数量的规律来划分，可以初步划分为以下3个阶段。

第一阶段：2002—2006年（初始阶段），这一阶段发表文献数量相对较少，为全部样本文献的25%，但是引用频次非常高（占样本文献全部被引次数的48.6%），多为本领域研究的概念界定、构念形成的研究，内容上主要为定义分析、机制探讨、重要性解释、国外理论与经验的引入等。

第二阶段：2007—2009年（快速增长阶段），这一阶段发表文献数量最多，为全部样本文献的48%，但是引用次数较低，仅为37.1%。这一阶段

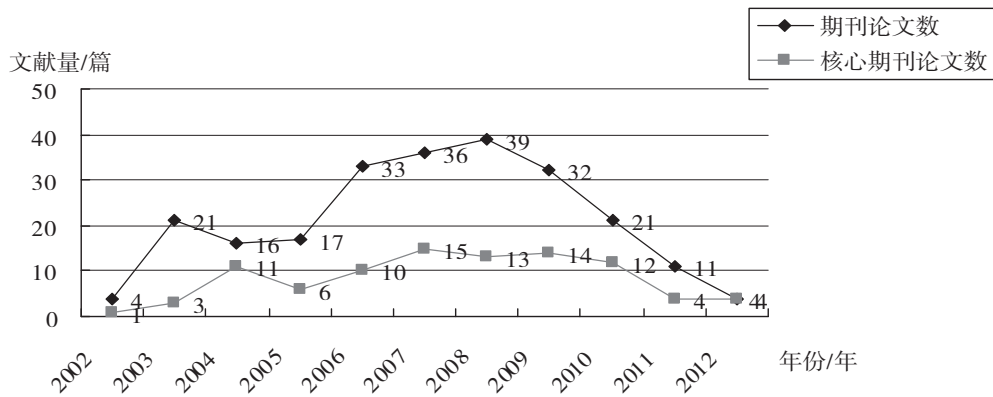


图1 国内科学数据共享领域期刊文献的年度

明显表现出具体构建或者应用科学数据共享的策略、对策措施、系统建设、环境因素分析、制度保障等。

第三阶段：2010年以后（转变阶段），这一阶段虽然发表的论文数量较少，但是核心论文数与全部论文数的差距缩小，一定程度上反映研究的质化发展，更加注重研究的品质，发表的期刊级别也比以往的要好，同时研究的内容更加侧重于具体的共享平台建设与计算机网络、数据库建设、元数据处理等关键性技术的实现，应用的可操作性与对策指导的可实现性日益完善。

3.2 期刊来源分布

研究期刊的来源可以较为全面地了解现有研究领域的空间分布与学科特点。通过对样本文献的梳理，发现样本文献的期刊来源涉及143个期刊，核心论文发表在包括了EI、CSCD（核心与扩展）、CSSCI（核心与扩展）、中文核心在内的41种核心期刊上，涉及了综合科技类、管理类、情报图书类、法律类、人文类以及基础自然科学、物理学、化学、地理学、天文学、水利学等多个学科，如表1和图2所示。

通过表1可以看出，样本文献的期刊来源主要集中在综合科技类的《中国基础科学》、《中国科技资源导刊》、自然科学类的《国际地震研究》、《应用气象学报》等16种期刊上，其中，发表论文数超过2次以上的有39种期刊，占总期刊比例的27%，而超过4次以上的不到8%。所以，国内科学数据共享研究的期刊分布比较分散，虽然形成若干重点关注的期刊，但是总体而言并不集中。另外，从图2中可以得出，仅从期刊归属的学科类型分辨，自然

科学类（包括物理、化学、地理、水利等学科）以及综合自然科学类（各大学的自然科学版的学报）所组成的比例最大，所占超过40%，同时综合科技类、图书情报类、管理类所占比例其次，而医学类、综合人文类、法律类、经济类所占比例很小。其中，工业技术类期刊刊载的论文主要为2008年以后。另外，样本文献中刊载论文影响力最高的期刊分别为自然科学类的《水利学报》（EI源期刊）、管理类的《科学学与科学技术管理》（国家自然科学基金委指定的A类期刊）、情报类的《图书情报工作》，而这三种期刊都不是各自领域的顶级期刊，除了《中国基础科学》、《中国科技资源导刊》等刊载了较多的相关论文外，其他期刊的论文数多未达到

表1 样本文献的期刊分布（部分）

期刊名称	载文数量/篇	占文献比例
中国基础科学	25	10.7%
中国科技资源导刊	9	3.8%
国际地震动态	6	2.6%
应用气象学报	5	2.1%
科技管理研究	5	2.1%
太原科技	5	2.1%
地球科学进展	4	1.7%
图书馆论坛	4	1.7%
中国卫生工程学	4	1.7%
现代情报	4	1.7%
地球信息科学	4	1.7%
情报杂志	3	1.3%
图书情报工作	3	1.3%
地球信息科学学报	3	1.3%
中国科技论坛	3	1.3%
水利学报	3	1.3%

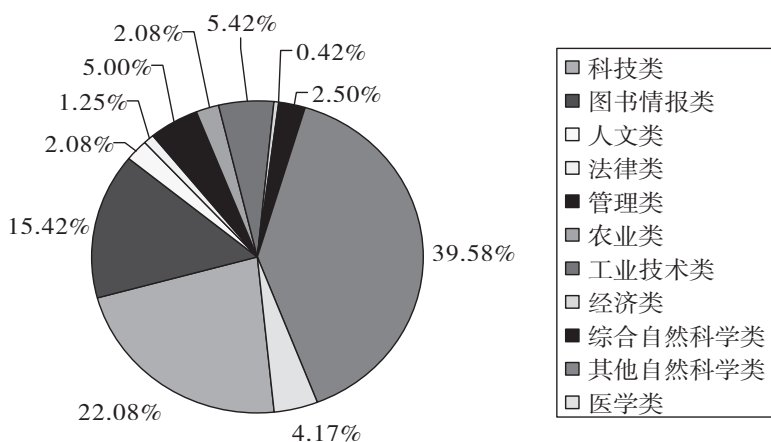


图2 样本文献的期刊类型分布

5篇以上，其中非核心期刊所占的比重较大。

4 核心作者与研究机构

4.1 核心作者分布

核心作者是领域研究的专家，通过了解核心作者群可以更好地掌握领域研究的前沿态势与重要理论。样本文献一共涉及451位作者（个别重名作者已做区分），其中独立撰写的有68位作者。本文参照储节旺、孙晓宁构建的核心作者评价体系进行筛选^[33]，依据普莱斯定律可以推算出本研究领域的核心作者发表的论文最低数为3，利用SATI3.2软件的频次统计功能得到42位核心作者的候选人，占样本全部作者比例的9%。为了弥补单一参照论文数量的评价标准，本文引入了对候选者论文数量与被引频次的加权赋值，并按照如下的规则进行统计：每篇论文计1分，独立作者权重为1，两位合作的，第一作者为0.6，第二作者为0.4；三人以上（包括三人）合作完成的，第一作者为0.6，第二作者为0.3，其他作者为0.1，然后将计算出的分数乘以一个系数（学者在样本文献中被引总次数/样本文献中第一被引用学者的总次数），最后再加上原来的分数，规定超过1.5分以上的为核心作者，最终得到校正后的结果，如表2所示。

通过表2可以看出，该领域研究的核心作者主要集中在中国科学院、中国农业科学院、国家气象局以及中国医学科学院等机构的学者，依照规则统计出的排名显示：孙九林名列该领域研究的第一位（孙九林是中国工程院院士、中国科学院研究员、

博士生导师，是资源信息领域的权威学者）。其他排名较高的核心作者如诸云强、王卷乐、刘闯、朱星明、张莉也都是领域内的知名学者。另外，核心作者的排名与引用次数的排名基本相似，也反映出引用频次指标的重要价值。但是，核心作者们主要是自然科学领域的研究者，而社会人文类或工程技术类的学者并不多见，且论文数超过4次以上的学者比例仍然较少，一定程度上反映了领域研究的核心作者群尚未完全形成，其他学科背景的学者涉入也较少，多元化交叉研究的趋势不够明显。为了凸显不同研究机构的不同学者之间相互合作关系，笔者利用样本文献中发表次数超过1次的作者的合作频次与单位关联情况，应用SATI3.2软件构建作者共现的二值矩阵，并利用ucinet6软件输出共现网络图谱^[34]，如图3所示。

图3中深颜色的点为高中心度的网络节点。由图3可知样本文献中存在较为集中的作者合作关系，分别形成2个合作较为紧密的作者群，第一个为李学良、王建国、李志雄等为主要节点的关系群，第二个是以孙九林、王卷乐、刘润达等为主要节点的关系群。两个关系群中绝大多数作者的单位分别为：国家地震相关研究机构与中国科学院所属各科研机构。由此可见，国内该领域的研究已经初步形成一定规模的学者合作研究的集群，但是数量有限，另外命题合作研究的趋势也日益凸显（合作完成论文所占样本文献的比重超过70%）。

4.2 研究机构分布

为了便于分析研究机构的分布，本文只对样本

文献中的第一作者单位进行统计,并将不同研究机构下的不同部门统一命名为机构名称,不再单独分类统计。经过统计,一共整理出样本文献所涉及的34个科研机构,如表3所示。

由表3中的数据可以看出,国内科学数据共享研究机构的分布有一定集中趋势,发表论文数超过3篇的研究机构有15个,这些机构总的发表论文数占样本文献总数的61%。其中,发表文献数最多的

几个依次是中国科学院(所占比例10%)、中国农业科学院(比例6%)、中国气象局(比例5%)、中国地震局(5%)。总体看来,对于该命题研究主要以中国科学院以及政府专业部门所属的各研究机构为主,高等院校参与研究的程度不高。另外,从研究机构被引频次的比较,中国科学院的文献被引总次数为543次,占全部被引总次数的32%,而中国农业科学院以及中国气象局、地震局等所占比例没有

表2 样本文献的核心作者排名(部分)

排名	作者	发文量	工作单位	被引次数	得分
1	孙九林	10	中国科学院地理科学与资源研究所	159	8.80
2	诸云强	9	中国科学院地理科学与资源研究所	82	7.12
3	王卷乐	8	中国科学院地理科学与资源研究所	108	5.71
4	刘闯	4	中国科学院地理科学与资源研究所	111	5.09
5	朱星明	6	中国水利水电科学研究院	86	4.78
6	张莉	4	中国农业科学院农业信息研究所	12	4.30
7	王国复	4	中国气象局国家气象信息中心	87	3.71
8	路鹏	6	防灾科技学院	26	3.49
9	李集明	6	中国气象局国家气象信息中心	109	3.37
10	张英俊	4	太原科技大学计算机科学技术学院	19	3.25
11	刘润达	5	中国科学院地理科学与资源研究所	6	3.11
12	樊秀娥	5	中国疾病预防控制中心	16	3.08
13	廖顺宝	4	中国科学院地理科学与资源研究所	41	2.89
14	刘德培	3	中国医学科学院	8	2.31
15	赵红	4	中国医学科学院	7	1.98
16	王巧玲	3	中国科学院国家科学图书馆(武汉)	9	1.90
17	尹岭	7	解放军总医院	16	1.87
18	张小红	3	中国农业科学院农业信息研究所	17	1.77
19	徐枫	3	国家信息中心	79	1.76
20	刘润达	5	中国科学技术信息研究所	72	1.75

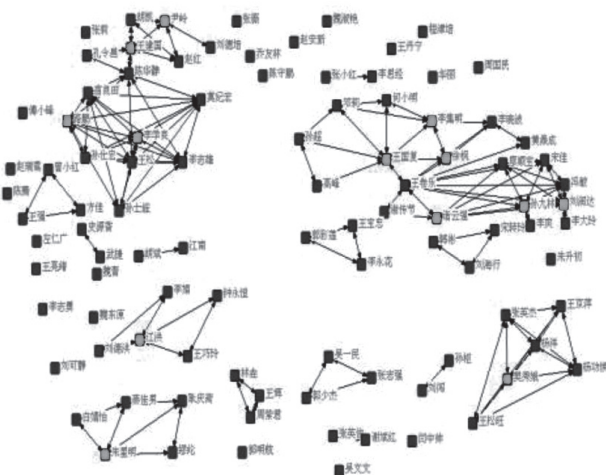


图3 样本文献作者合作关系的网络图谱

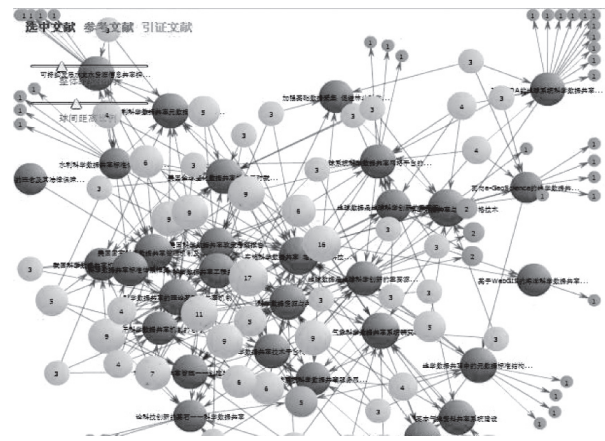


图4 样本文献中的高被引文献组的互引图

表3 样本文献研究机构的分布(部分)

排名	研究机构	发文章
1	中国科学院	51
2	中国农业科学院	14
3	中国气象局	11
4	中国地震局	10
5	中国科学技术部	7
6	中国水利水电科学研究院	7
7	防灾科技学院	6
8	解放军总医院	6
9	中国医学科学院	6
10	中国科学技术信息研究所	5
11	中国疾病预防控制中心	5
12	太原科技大学	4
13	华南理工大学	3
14	国家海洋局	3
15	国家信息中心	3

超过3%,由此可以看出中国科学院在该领域研究中的贡献。

5 被引文献, 研究主题及其频次统计

5.1 高被引文献

论文的被引频次能够反映出文章的学术价值,所以按照论文被引频次高低排列,从样本文献中选择出被引频次大于20的学术论文,如表4所示,同时为了分析出样本文献中高被引文献之间的相互引证关系,本文利用CNKI文献分析系统^[35],输出文献互引图,如图4所示。

由表4可以看出,国内科学数据共享研究的高被引论文多发表在2005年以前,引用频次基本保持在25~55次之间,没有很高的被引频次论文出现。另外,从表4中15篇论文的发表期刊来看,普通期刊《中国基础科学》发表论文数最多,有8篇论文,而其他7篇发表的期刊都为核心期刊。由图4中可以看出,徐冠华院士(原科技部部长)的《实施科学数据共享,增强国家科技竞争力》以及孙九林院士的《科学数据资源与共享》以及科学数据共享调研组的《科学数据共享工程的总体框架》在样本组中被引次数最多,一定程度上反映出几位学者在本研究领域的重要地位。

5.2 研究主题分析

学术论文中的高频关键词可以反映出某个学科领域的主题概况以及研究热点。在进行分析之前,本文对SATI3.2统计得出的关键词分布表进行预处理:(1)合并同义词,例如科学数据的共享与科学数据共享是同义词,合并为一个;(2)删除一些与研究无关的人名、职称、地名,例如张先、副部长、黑龙江省等。经过预处理后,选择频次高于5次以上(包括5次)的作为高频关键词(表5),利用SATI3.2软件构建高频关键词的共现相似矩阵,因为共现相似矩阵中的0值比较多,需要转化为相异矩阵,将相似矩阵中的数值用1相减,并把转化后的相异矩阵导入SPSS18.0软件进行层次聚类分析^[36],输出结果如图5所示。

通过表5可以看出,科学数据共享是频次最高

表4 样本文献中高被引文献统计

排名	题名	年份	被引频次
1	实施科学数据共享 增强国家科技竞争力	2003	55
2	地球科学数据共享与数据网格技术	2002	45
3	美国国有科学数据共享管理机制及对我国的启示	2003	43
4	气象科学数据共享系统研究综述	2004	42
5	科学数据共享的理论基础与共享机制	2003	32
6	科学数据资源与共享	2003	32
7	基本气象资料共享系统建设	2004	32
8	地学数据共享中的元数据标准结构分析与设计	2005	32
9	中国气象科学数据共享服务网总体设计与建设	2004	31
10	水利科学数据共享元数据理论的应用探讨	2005	29
11	科学数据共享工程的总体框架	2003	27
12	美国全球变化数据共享的经历对我国数据共享决策的启示	2002	27
13	科学数据的共享管理——创建共享新秩序	2003	26
14	地球数据是地球科学创新的重要源泉	2003	26
15	科学数据共享标准体系框架	2003	26

的关键词,这也与本文检索的主题词有密切联系,对高频关键词进行归类,可以初步划分为科学数据共享相关的行为、标准、数据库、数据、服务、技术、理论6类。图5反映了国内研究的大致分布,可以分为:关键概念、各类数据、数据库、服务、共享平台、共享网络、法律政策6个类别。结合上述分析结果,同时参照国内相关文献综述的内容^[22],本文将国内科学数据共享的相关研究划分为5个主题范围,包括:基本理论研究、数据源与数据库、信息共享与共享网络、共享平台与服务、标准规范与政策法律。

基本理论研究主要包括了科学数据共享的基本概念、意义或者重要性、价值显现以及共享过程中遇到的挑战、共享研究的范式等,也包括了共享机制的讨论、国内外相关理论的引入与经验的借鉴^[37]。这一主题范围的工作是本领域基础性研究,在早期阶段所占比例较高,且被引次数也是很高,随着研究的不断深入,该部分的研究将会日益完善,最终形成中国情境下的理论体系。

数据源与数据库的研究则包括了科学数据的识别与采集、数据编码与加工、数据存储与管理、各类科学数据库的开发以及其他相关技术性研究等^[38]。科学数据的共享离不开系统性的科学数据存储、加工与整合,高效的数据库管理与先进的数据加工技术可以提高数据编码、传输的效率,深入挖掘有价值的数据库,积极推动知识的扩散。

信息共享与共享网络的研究重点集中于信息或数据传递、共享模式、互联网技术应用与发展等,主要目的是为了便于信息接收者如何更好、更快地从数据库中获取有价值的信息,从而进行有效的整合^[39]。另外,不同信息所有者之间如何构建一个有效的联通网络,实现信息的互动与共享也是其研究的内容。

共享平台与服务的研究主要涉及科学数据共享平台的建设、平台的服务模式与服务能力、平台的管理与协调、开放式数据提供与更新等^[26]。提倡科学数据共享平台是为了克服数据分散、信息不对称以及实际利用率低的缺陷,从而更好地推动生产力的进步。

标准规范与政策法律则包括了本领域重要的制度标准与技术规范的确立、国家宏观政策的建议与措施、信息立法保护与知识产权研究等内容^[40]。科学数据共享的实现不仅需要硬件的支持,更需要相关软件的保障,例如技术领域的标准规范有利于信息传递与接收,而专项法律的颁布有利于保障共享过程中各方的正当利益,国家政策的出台有利于配套资金的支持、科研课题的资助研究等。

5.3 主题频次统计

笔者以聚类分析得到的主题分布为依据,对234篇样本文献进行主题频次的统计,结果如表6所示。

由表6所示,在5个主题类别中,共享平台与

表5 样本文献的高频关键词统计

序号	关键词	频次	序号	关键词	频次
1	科学数据共享	73	17	数据集	8
2	数据共享	66	18	Web	7
3	科学数据	53	19	数据共享服务	7
4	元数据	35	20	数据管理	7
5	共享	22	21	标准规范	7
6	数据库	19	22	农业科学数据	7
7	共享平台	17	23	科技部	6
8	数据资源	14	24	科技资源	6
9	共享工程	12	25	科技创新	5
10	数据中心	10	26	数据共享平台	5
11	共享服务	9	27	WebGIS	5
12	共享管理	9	28	目录服务	5
13	地球系统科学	8	29	气象科学数据	5
14	信息共享	8	30	科学数据中心	5
15	科学数据管理	8	31	农业	5
16	共享机制	8	32	XML	5

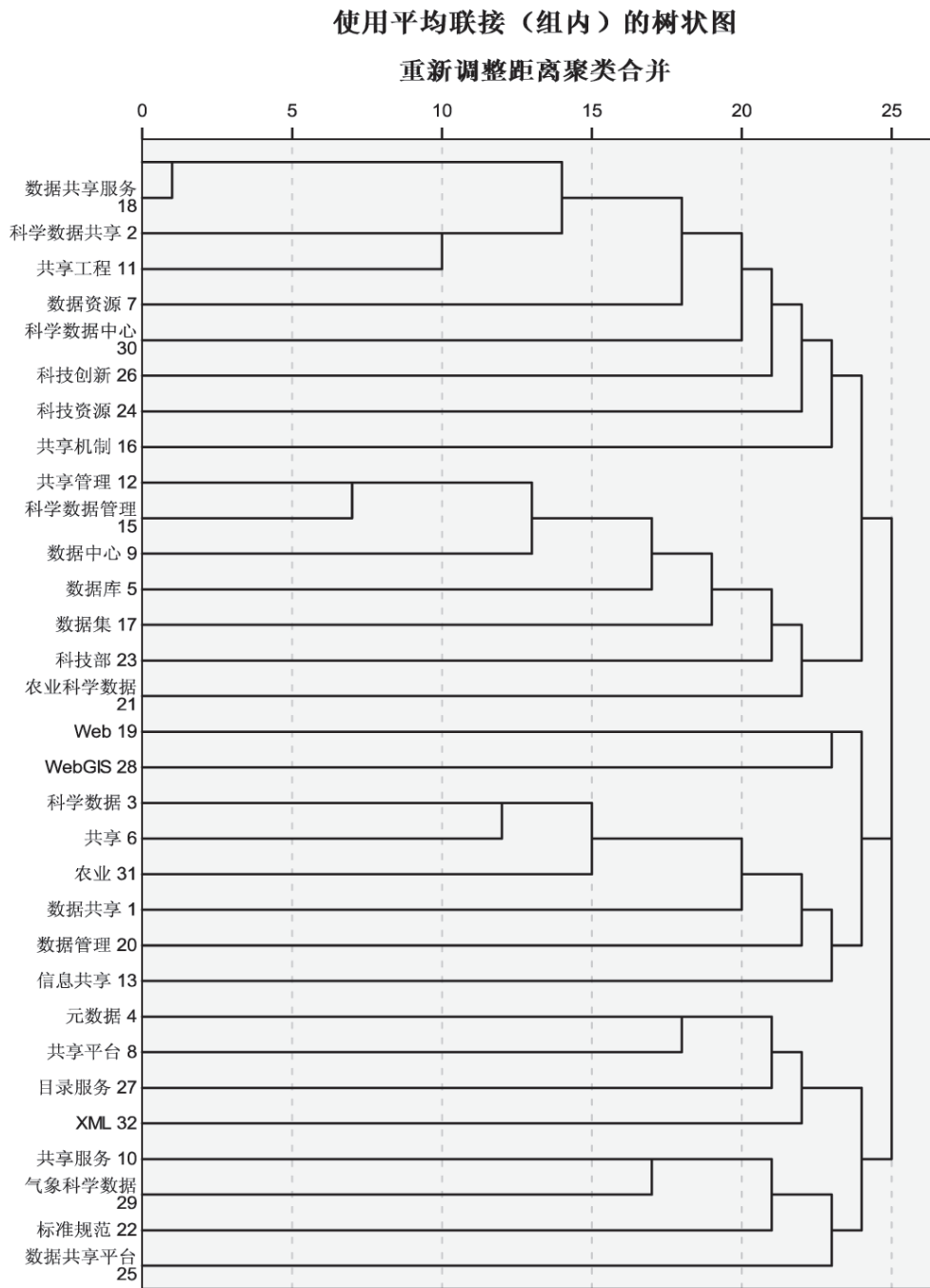


图5 样本高频关键词聚类树状图

服务模式的频次最多，占全部样本文献的36%，超过1/3的比重，同时看不同时期的频次分布，该命题的研究仍然有上升趋势。另外，基本理论研究、信息共享与共享网络、标准规范与政策法律的比重基本相同，都在13%左右，而数据源与数据库的比重第二。未纳入以上5个类别的文献共有14篇，主要是科技进展报道、现状调查分析、政策考察报告等内容，所以单独放到其他类型中进行统计。从近

3年（2010—2012）的频次分布来看，共享平台与服务模式仍是研究的焦点，数据源与数据库主题的频次上升，主要是因为各项先进技术的出现、元数据应用的普及以及新环境——“大数据时代”的来临，这些对海量科学数据的编码、加工与转换以及数据存储模式与管理机制提出了新的要求。而各项标准规范及政策法律的讨论，例如法律制度、行业技术标准、政府宏观策略等也是近期讨论较多的内容。

表6 样本文献的主题频次统计

时期\主题	科学数据共享 基本理论研究	数据源与 数据库	信息共享与 共享网络	共享平台与 服务模式	标准规范与 政策法律	其他类型
2002-2006年	12	15	11	20	9	3
2007-2009年	11	13	12	31	13	6
2010-2012年	7	14	8	33	11	5
总计次数	30	42	31	84	33	14
总百分比	13%	18%	13%	36%	14%	6%

6 科研基金资助情况

科研基金的资助是高质量论文产出的重要保障,也是一项重要的科研评价标准。通过对样本文献的分析,其中获得科研资助的有65篇,占全部样本文献的28%,并对各类科研基金进行分类统计,如表7所示。

从表7可以看出,受资助的项目主要来自于国家级课题基金,例如国家科技基础条件平台建设项目(占全部基金资助文献比例的80%)、国家科技基础性专项计划项目(所占比例为10%)以及国家自然科学基金(所占比例为15%)等。另外,表中的数据也反映了省级科研资助项目较少,仅有山西省、河北省、上海市等少数省份的课题资助。仅对该领域研究而言,地方财政支持的力度略显不足。

7 结论

(1)国内相关研究主要兴起于21世纪以后,文献年度变化趋势总体呈现出先增后减的波动趋势,2007-2009年期间是该命题研究的高峰阶段。

(2)期刊来源有一定集中趋势,特别是以《中

国基础科学》与《中国科技资源导刊》两类综合性科技期刊发表论文数最多。

(3)研究领域已经初步形成一定的核心作者群,孙九林、诸云强等学者是本领域的权威人士。

(4)从研究机构的分布来看,中国科学院是相关研究机构中发表论文数最多的机构。

(5)论文的高被引频次主要集中在25~55次之间,发表的相关文献多在2005年以前,处于研究的早期阶段。

(6)受课题资助的论文仅占全部样本文献的28%,主要以国家级科研课题资助为主,省级课题资助很少。

(7)通过高频关键词的聚类分析,研究主题的范围可以划分为基本理论研究、数据源与数据库、信息共享与共享网络、共享平台与服务、标准规范与政策法律5个方面。

(8)从研究主题的频次统计可知,共享平台与服务是现有命题研究的焦点,占到全部样本文献的36%,其次是科学数据的数据源及数据库方面的研究,比例为18%,其他3个方面主题的频次基本相同。

未来,伴随着“大数据时代”的到来、知识管理应用的普及,国内外对元数据与科学数据共享的研究以及知识管理机制的建立会更加深入,通过云处理能力的提升、共享网络的优化、共享机制的健全、共享平台的服务化,真正推动科学数据共享事业的进步^[41]。

另外,本文将文献计量分析的结果同王巧玲、钟永恒与江洪三位学者共同发表在《情报杂志》上有关2007年以前该领域研究现状的结论^[42]进行比较,可以发现:(1)学者合作发表的趋势日益明显,发表2篇、3篇以及4篇以上的人数都有明显增加;(2)《中国科技资源导刊》自2008年创办以来,经过5年的时间,已经成为领域内发表论文数第二的

表7 样本文献的科研基金资助情况统计

序号	科研基金名称	文献量
1	国家科技基础条件平台建设项目	52
2	国家自然科学基金	10
3	国家科技基础性专项计划项目	6
4	中国科学院知识创新工作项目	3
5	国家软科学研究项目	2
6	国家高新技术研究发展计划项目	2
7	中国科协调研课题	1
8	山西省自然科学基金项目	1
9	河北省软科学研究项目	1
10	上海科技发展基金	1

期刊,另外核心期刊的比重相比2007年以前日益提高;(3)图书情报类的期刊比重日益增加,特别是近5年相关论文数不断增加,反映了多学科交叉研究的倾向。

本文的研究仍有一定的局限。首先,样本选择只考虑CNKI收录的期刊,有部分期刊尚未收录,部分国内学者的研究成果发表在国外期刊上,也未纳入样本文献中。另外,本研究只聚焦于国内,没有借鉴国外研究的最新进展与态势进行相应的反思。

参考文献

- [1] 王培正,张志强,吴一鸣.科学数据共享方式研究[J].科技管理研究,2010(17):201-204.
- [2] 赵卫利,陈晓毅,靳红.科学数据共享平台,支撑优势产业发展研究[J].科技与经济,2008(2):53-55.
- [3] 徐冠华,实施科学数据共享 增强国家科技竞争力[J].中国基础科学,2003(1):7-11.
- [4] 李娟,刘德洪,江洪.国际科学数据共享原则和政策研究[J].图书情报工作,2008(12):77-80.
- [5] 王巧玲,钟永恒,江洪.英国科学数据共享政策法规研究[J].图书馆杂志,2009(10):62-65.
- [6] 黄鼎成,李晓波,莫纪宏.科学数据共享法规体系建设的若干思考[J].中国基础科学,2003(6):40-45.
- [7] 王倩,王延政,杨虹.科学数据共享[J].农业网络信息,2008(2):29-30.
- [8] 朱星明,白婧怡,蔡佳男.水利科学数据共享体系建设初探[J].中国水利,2006(5):47-48,60.
- [9] 刘德培,尹岭.中国医药卫生科学数据共享实践[J].计算机应用与软件,2006(10):1-3,28.
- [10] 王广华.国土资源科学数据共享研究综述[J].测绘通报,2007(4):34-37.
- [11] 刘瑞丰,蔡晋安,彭克银,等.地震科学数据共享工程[J].地震,2007(2):9-16.
- [12] 安基文,庄大方,袁文.面向地学计算的资源环境科学数据共享平台的设计[J].地球信息科学,2007(3):34-39.
- [13] 李纪宾,邹小农,乔友林.元数据及其在医学科学数据共享中的应用[J].医学情报工作,2006(1):10-14.
- [14] 王卷乐,诸云强,谢传节.地球系统科学数据共享网络平台的设计和开发[J].地学前缘,2006(3):54-59.
- [15] 陈燕,吕晓春,孙静涛.焊接科学数据共享体系设计与元数据研究[J].焊接,2006(6):56-60.
- [16] 李雪山,熊燕.基于元数据的交通科学数据共享平台的设计与实现[J].铁路计算机应用,2007(1):24-26.
- [17] 张小红,李思经.农业科学数据共享服务研究[J].中国科技论坛,2006(5):127-130.
- [18] 陈腾,张子平.国土资源科学数据共享平台的框架研究与设计[J].山西建筑,2007(6):360-361.
- [19] 黄鼎成,李晓波,王卷乐.浅谈科学数据共享工程建设的战略取向[J].中国基础科学,2005(5):29-35.
- [20] 陈军,王春卿.关于科学数据共享机制的思考[J].中国基础科学,2003(1):42-45.
- [21] 朱雪忠,徐先东.浅析我国科学数据共享与知识产权保护冲突与协调[J].管理学报,2007(4):477-482,487.
- [22] 路鹏,苗良田,莫纪宏,等.我国科学数据共享现状[J].国际地震动态,2007(6):26-32.
- [23] 朱星明,耿庆斋,蔡佳男.水利科学数据共享的现状与发展趋势[J].中国水利,2008(14):47-50.
- [24] 孙晨晔,朱德米.基于网络治理的长三角地区先进制造业科学数据共享机制的构建[J].上海管理科学,2008(5):64-69.
- [25] 刘闯.美国国有科学数据共享管理机制及对我国的启示[J].中国基础科学,2003(1):36-41.
- [26] 胡海燕,刘世洪.论“国家农业科学数据共享平台”的内容与服务[J].农业图书情报学刊,2005(2):214-217.
- [27] 赵胜钢.国家农业科学数据共享平台体系结构研究[D].北京:中国农业科学院,2009:33-35.
- [28] 申丹娜.美国实施全球变化研究计划的协作机制及其启示[J].气候变化研究进展,2011(6):449-454.
- [29] 李思经,张小红.农业科学数据共享管理研究[J].科技导报,2006(8):80-82.
- [30] 马海龙,陈佑启,邹金秋.“国家农业科学数据共享中心农业区划数据分中心”系统设计与建设[J].中国农业资源与区划,2010(1):75-80.
- [31] 李莉.疾病预防控制科学数据共享与实证研究[J].预防医学情报杂志,2011(1):53-55.
- [32] 武士华.科学数据共享与科技情报工作[J].中华医学图书情报杂志,2005(5):1-3.
- [33] 储节旺,孙晓宁.国内图书馆知识管理研究现状的文献计量分析[J].图书馆理论与实践,2012(9):21-26.
- [34] 刘启元,叶鹰.文献题录信息挖掘技术方法及其软件SATI的实现——以中外图书情报学为例[J].信息资源管理学报,2012(1):50-58.
- [35] 黄远辉.基于综述文献和CNKI“知网节”的科技查新[J].情报科学,2011(2):258-260,263.
- [36] 储节旺.国内图书情报领域知识管理研究方法的共词分析[J].情报杂志,2012(11):45-50.
- [37] 严冬梅,尚翔.论科技创新的基石——科学数据共享

- [J]. 科学管理研究, 2005(1):20-22.
- [38] 张雷, 何险峰, 陆冬云, 等. 基于 Web Services 技术的科学数据共享研究[J]. 计算机与应用化学, 2005(2):52-56.
- [39] 胡斌, 江南, 赵钢, 等. 基于语义 P2P 的地理共享网络中地理服务发现机制[J]. 地理信息科学, 2008(2):177-182.
- [40] 路鹏, 苗良田, 莫纪宏, 等. 构建完善合理的科学数据共享政策法规体系[J]. 国际地震动态, 2008(3):25-33.
- [41] Demchenko Yuri, Zhao Zhiming, Grosso Paola, et al. 科研信息化基础设施的大数据挑战[J]. 中国科技资源导刊, 2013(1):30-35.
- [42] 王巧玲, 钟永恒, 江洪. 我国科学数据共享研究的文献计量分析[J]. 情报杂志, 2008(7):128-130, 134.

(上接第 26 页)

重要特征之一就是内在结构性, 学科间的“树形”体系结构可以作为组织科技信息资源的框架。按照该体系进行资源组织最大的优点是能够让使用者从全面宏观的角度把握科技信息资源, 而且能够很好地查看到最新的数据以及这些数据的历史发展脉络。同时, 在信息的查询界面上更加突出其友好性一面, 对科技信息资源进行“树形”的可视化展示, 并辅以人机对话式的检索窗口, 真正做到直观性和交互性的统一。

5 总结

在介绍云计算和科技信息资源概念的基础上, 分析了两者具有的特征, 试图在两者的特征之间找到内在的适应性。通过分析, 云计算可以为信息资源共享提供更为广阔的舞台, 因为云计算所具有的特征在某种意义上来说能够很好地适应科技信息资源共享的需求, 然后提出了对云计算环境下科技信息资源共享的几点思考, 包括云计算下科技信息资源的共享模式、资源的组织框架和运行机制。然而, 云计算环境下科技信息资源共享服务体系的真正实现是一个庞大的系统工程,

不仅需要技术层面的支持, 而且需要政府部门的支持以及配套的法律制度等提供保障。只有各个科技信息资源共享主体之间不断努力奋斗才能打造出更加便捷、完善的科技信息资源共享服务体系, 只有调动不同主体的积极性才能更好地促进云环境下的科技信息资源共享。

参考文献

- [1] Michael Stephens. Tame the Web [EB/OL]. [2012-05-24]. <http://tametheweb.com/about-michael-stephens/>.
- [2] Rajkumar Buyya, Chee Shin Yeo, Srikumar Venugopal, et al. Cloud Computing and Emerging IT Plat—rotors: Vision, Hype, and Reality for Delivering Computing as the 5th Utility[J]. Future Generation Computer Systems, 2009, 25(6):599-616.
- [3] 穆向阳. 云计算背景下跨媒体信息素养和认知模式的关系[J]. 情报杂志. 2012(3): 174-179.
- [4] 穆向阳, 缪宁, 陈明, 等. 云计算环境下 BI 对企业核心竞争力的影响[J]. 情报杂志. 2010(6):50-53.
- [5] 吴长旻. 浅析“科技资源共享”[J]. 科技管理研究, 2007(1):49-51.
- [6] 黄菁, 杨帆. Wiki 知识共享与企业 Wiki 理论初探[J]. 图书与情报, 2009(1):55-60.