

圖書館學 情報學 理論與實踐系列叢書

辛希五題

名家視點
Master viewpoint
第8輯

数据管理的研究与实践

SHUJU GUANLI DE YANJIU YU SHIJIAN



《图书情报工作》杂志社 编

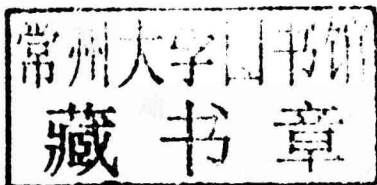


海军出版社

名家视点 第8辑

数据管理的研究与实践

《图书情报工作》杂志社 编



海洋出版社

2018年·北京

图书在版编目 (CIP) 数据

数据管理的研究与实践/《图书情报工作》杂志社编. —北京:海洋出版社, 2018. 1

(名家视点. 第8辑)

ISBN 978-7-5210-0015-3

I. ①数… II. ①图… III. ①数据管理-研究 IV. ①TP274

中国版本图书馆 CIP 数据核字 (2017) 第 330694 号

责任编辑:杨海萍 张 欣

责任印制:赵麟苏

海洋出版社 出版发行

<http://www.oceanpress.com.cn>

北京市海淀区大慧寺路8号 邮编:100081

北京朝阳印刷厂有限责任公司印刷 新华书店北京发行所经销

2018年4月第1版 2018年4月第1次印刷

开本:787 mm×1092 mm 1/16 印张:23.5

字数:410千字 定价:52.00元

发行部:62132549 邮购部:68038093 总编室:62114335

海洋版图书印、装错误可随时退换

《名家视点丛书》编委会

主 任：初景利

委 员：杜杏叶 易 飞 徐 健 王传清

王善军 刘远颖 赵 芳 谢梦竹

胡 芳 栾瑞英 盛怡瑾 袁贺菊

王 瑜

序

伴随着“狗年”的来临，由《图书情报工作》杂志社策划编辑、海洋出版社正式出版的《名家视点：图书馆学情报学档案学理论与实践系列丛书》第8辑如约而至，就要与广大读者见面了。这也是《图书情报工作》杂志社和海洋出版社联袂在狗年为广大的读者献上的一份小小的礼物。

本辑丛书包括四本书：《阅读推广的进展与创新》《面向 MOOC 的图书馆嵌入式服务创新》《数据管理的研究与实践》《智慧城市与智慧图书馆》。四本书所有文章均是从《图书情报工作》近些年所发表的优秀论文中遴选出来的。可以说，这四个主题都是当下学界业界所关注的热点或前沿领域，是图书馆学情报学理论与实践的新发展，也是国内近些年关于这些领域研究成果的集中体现。

《阅读推广的进展与创新》共计收录 29 篇文章。阅读推广是图书馆的一种重要服务模式，既是图书馆馆藏资源宣传推广的一种策略，也是拉近图书馆及其馆藏与读者之间距离的一种重要手段，更是提升公众文化素质与阅读素养的一种重要机制。从学术的角度，阅读推广的研究主题并不是创新，但实践上的异常活跃给阅读推广研究带来了新的生机与活力。本专辑的内容不仅展现了关于阅读推广的若干基本理论研究成果和多个国家阅读推广的实践经验，还重点汇集了多个图书馆在阅读推广方面的成功案例，值得学习和借鉴。

《面向 MOOC 的图书馆嵌入式服务创新》收录 27 篇文章，分“理论篇”“建设篇”“服务篇”“综述篇”四部分，阐述了图书馆的环境下 MOOC 的应用与发展。MOOC 在图书馆中的引入和应用已有数年的历史，但其意义和价值仍待不断地开发，其应用前景非常乐观。MOOC 以其独特的教学模式深刻地影响了大学教育，也为图

书馆创新服务提供了新的手段和契机。国内外图书馆在 MOOC 教学与服务方面已经有了不少的探索。本书可以说是从一个侧面反映了这些探索所取得的成果。

《数据管理的研究与实践》共收录 27 篇文章，分“理论篇”“国外篇”“国内篇”，一定程度上客观总结了国内外在数据管理的研究与实践方面所取得的最新进展。数据管理（或称科研数据管理、科学数据管理、数据监护等）是数据密集型科研范式（第四范式）转变的必然要求，也为图书馆信息服务、知识服务从基于文献到基于数据提供了新的机遇与新的能力。但总体而言，对国内的图书情报工作而言，数据管理还是新生事物，我们对它的认识与应用的能力还非常有限。本书所介绍的相关内容对于我们更好地理解数据管理，推动数据管理融入图书馆业务体系，建立数据管理平台与服务能力，都是很有启发价值的，特别是国内外图书馆在数据管理方面的一些探索，表明数据管理已经不是概念层面的问题，而是在实践中已经有了长足的发展。

《智慧城市与智慧图书馆》共收录论文 26 篇。“智慧”是一个非常时髦的词汇。智能技术的发展与应用，使得“智慧城市”“智慧社区”“智慧校园”乃至“智慧地球”成为可能。可以说，智能无处不在，智慧无所不能。同样，如果城市是智慧的，校园是智慧的，图书馆如果还不是智慧的，那图书馆是否还有存在的必要？因此，加快智慧图书馆的建设绝不是口号和噱头，而是当务之急，具有迫切的需求。2017 年，国内对智慧图书馆的讨论异常热烈，许多会议都将智慧图书馆列入探讨主题，许多期刊发表了许多篇智慧图书馆的文章。如果说将 2017 年定为“智慧图书馆元年”，也不为过。本书将为智慧图书馆的研究与实践提供助推器，希望国内图书馆更多地关注智慧图书馆，更多地参与智慧图书馆的建设，尽早实现智慧图书馆的目标。

《图书情报工作》至今已经走过 62 个年头，也处于其历史发展的最好时期。2017 年各项计量指标均名列前茅，而且还首次获得中国科学院科技期刊排行榜奖励，特别是首次获得“全国百强科技期刊”。杂志社不仅立足办好期刊，更快地发表更多的优秀成果，还

积极承担传播知识的社会责任，每年举办多场学术会议和培训。出版专辑也是这样一种责任的体现，使得分散的相关主题的研究成果得以通过图书的形式再次揭示与展现，推动所发表的成果的增值和再利用。

感谢收录本专辑的各篇论文的作者的贡献，感谢广大读者对本专辑和本刊多年来的关注、厚爱和支持。在许多人的观念里，图书情报是传统行业，但这一行业在需求与技术的双驱动下，正在焕发前所未有的青春。通过创新与变革，重新定位图书情报的专业角色，重新塑造图书情报的职业形象，重新构建图书情报的职业能力，是时代赋予我们这一代图情工作者的神圣责任。

祝大家狗年“旺，旺，旺”！

初景利

中国科学院大学经济与管理学院图书情报与档案管理系主任

《图书情报工作》杂志社社长、主编，教授，博士生导师

2018年2月9日 北京 中关村

目 次

理 论 篇

高校图书馆参与科学数据验证的前景分析	(3)
利益相关者视角下图书馆参与科学数据管理的分析	(14)
图书馆在科学数据管理中的角色定位研究	(24)
面向科研第四范式的科学数据监管体系研究	(33)
国内外高校科学数据管理和机制建设研究	(47)
高校科学数据管理服务实践研究及建议	(59)
基于科学数据管理的图书馆数据服务研究	(76)
国内外大学科学数据监管比较研究	(87)
对高校科学数据管理平台建设的建议	(99)

国 外 篇

英美数据管理计划与高校图书馆服务	(113)
英美社会科学数据管理与共享服务平台调查分析	(129)
英美政府数据开放平台数据管理功能的调查与分析	(149)
国外高校科学数据生命周期管理研究及借鉴	(164)
国外高校数据监管项目的调研与分析	(174)
英国科研资助机构的数据管理与共享政策调查及启示	(193)
英国基金机构数据管理计划的实践调查与分析	(205)
加州大学伯克利分校数据管理的实践剖析	(218)

美国社会科学数据管理联盟(Data-PASS)的发展与借鉴	(230)
国外大学数据监护教育的调查与分析	(240)

国内篇

高校研究数据管理需求调查实践探索	(257)
基于用户调查的高校科学数据管理需求分析	(278)
高校图书馆科学数据管理与服务初探	(291)
面向科研数据管理的高校学科馆员能力建设研究	(301)
社会科学数据管理服务平台系统选型研究	(316)
农业科学数据监管平台构建研究	(326)
基于 Dspace 构建高校科学数据管理平台	(337)
国内数据监护平台研究热点与进展探析	(346)

参考文献

(1)	
(2)	
(3)	
(4)	
(5)	
(6)	
(7)	
(8)	
(9)	
(10)	
(11)	
(12)	
(13)	
(14)	
(15)	
(16)	
(17)	
(18)	
(19)	
(20)	
(21)	
(22)	
(23)	
(24)	
(25)	
(26)	
(27)	
(28)	
(29)	
(30)	
(31)	
(32)	
(33)	
(34)	
(35)	
(36)	
(37)	
(38)	
(39)	
(40)	
(41)	
(42)	
(43)	
(44)	
(45)	
(46)	
(47)	
(48)	
(49)	
(50)	
(51)	
(52)	
(53)	
(54)	
(55)	
(56)	
(57)	
(58)	
(59)	
(60)	
(61)	
(62)	
(63)	
(64)	
(65)	
(66)	
(67)	
(68)	
(69)	
(70)	
(71)	
(72)	
(73)	
(74)	
(75)	
(76)	
(77)	
(78)	
(79)	
(80)	
(81)	
(82)	
(83)	
(84)	
(85)	
(86)	
(87)	
(88)	
(89)	
(90)	
(91)	
(92)	
(93)	
(94)	
(95)	
(96)	
(97)	
(98)	
(99)	
(100)	

高校图书馆参与科学数据验证的前景分析*

在大数据时代背景下,国家竞争焦点已从资本、人口、资源的竞争转向了数据竞争,世界各国越来越重视大数据。美国作为世界科技强国,已于2012年3月29日正式启动“Big Data Research and Development Initiative”计划^[1],正式将“大数据”提高到国家战略层面;我国也加快了大数据战略制定的步伐,2015年8月31日,国务院印发了《促进大数据发展行动纲要》的通知,提出了“全面推进我国大数据发展和应用,加快建设数据强国”的“决策部署”^[2]。科学数据作为大数据的一种,有更高的科学价值和社会价值,但是科学数据质量参差不齐、科学数据审查核验机制不够完善等,严重影响了其价值的发挥。另外,我国目前尚没有对科学数据进行验证的权威机构,高校图书馆作为科学研究活动的重要信息服务机构,有开展科学数据验证业务的良好历史机遇。本文拟对高校图书馆参与科学数据验证的必要性、可行性和参与途径进行研究,旨在为高校图书馆未来拓展科学数据验证业务提供参考。

1 高校图书馆参与科学数据验证的必要性

科学数据验证是指对科学研究最终产生的科学数据的完整性和真实性进行复查核验的过程,完整性是指数据在思想上不随时间推移而改变,保证数据整个生命周期的准确性和一致性,真实性是指这些数据是基于事实、真实可信的,科学数据验证对科学研究的后续进程有重要影响。

1.1 净化学术环境

随着 E-science 的发展,科学研究已进入第四范式,即数据密集型科学研究^[3]。科学研究产生的科学数据具有数量大、种类多、更为复杂的特点,在

* 本文系国家自然科学基金项目“开放数据下公共信息资源再利用体系的重构研究”(项目代号71373195)和武汉大学与中国科技信息研究所合作项目“科学文献的语义功能识别与深度利用”研究成果之一。

这种前提下,科学数据的完整性和真实性就显得十分重要。然而,现实生活中却时常发现科学数据造假的情况。如2002年“贝尔实验室科学家造假事件”,舍恩在《科学》、《自然》和《应用物理通讯》等全球著名学术期刊上发表了10余篇论文,而且涉及的都是超导、分子电路和分子晶体等前沿领域,但经过调查认定舍恩至少在16篇论文中进行了数据造假;2015年3月,英国大型学术医疗科学文献出版商现代生物出版集团(BioMed Central);因同行评议造假撤销了43篇论文;2015年5月,美国社会科学界爆出轰动性的丑闻:加州大学洛杉矶分校(UCLA)的政治系教授D.Green和博士生M.LaCour半年前在《科学》杂志发表的一篇广受关注的论文数据造假。

世界顶尖信息学院联盟iSchool主席M.Seadle教授在第二届iconference亚太地区分会(The Second Asian Pacific iConference)上做了关于“科研诚信”主题的报告,报告中提到,从2010年8月到2015年10月,学术界已发生学术剽窃案例310件,数据伪造案例249件,研究人员数据出错129件,这些使用并发表虚假数据的行为造成了学术资源浪费,严重破坏了公平、公正的学术环境,危害了科学研究的进程。因此,目前迫切需要成立科学数据验证的中介机构,对科学数据发表前的验证也将成为一种必然趋势。高校图书馆作为重要的科学研究信息服务机构,拥有大量的科学数据资源,应主动拓展业务范围,开展科学数据验证业务。科学数据验证可以大大减少科研人员伪造科学数据的机会成本,净化学术环境,营造公平、公正的学术科研竞争氛围,保证科学研究进程的顺利进行和科学数据的后续使用;同时能够促进国家基金的有效利用,减少学术资源和科研基金的浪费;还可以进一步保护科研人员科研成果的知识产权,防止其科学数据被他人剽窃或抄袭事件的发生,提高科学研究的效率。

1.2 提高图书馆地位

《科学》杂志主编艾伯茨说,学术不端是天大的事;中国工程院院士、中国中医科学院院长曾撰文写道:科学研究应首重诚信;中国工程院院士郑哲敏提出:像保护生命一样呵护科学诚信。一个合理的科学研究过程要求必须具有可重复性,因此,科学数据也具有可重复性的特点^[4],为了避免学术不端行为,需要对科学数据进行验证:一方面,能够使科学研究人员重新反思自己的学术伦理和社会责任,端正学术态度并自觉接受社会监督;另一方面,也能保障并加快对于科学数据的分享和转化利用,提升科学数据的社会价值。高校图书馆应在已有的业务基础上,开展科学数据验证的新业务。

目前,国内外越来越多的高校图书馆都积极参与科学数据管理实践中,

提供科学数据管理服务。英国数据存档(UK Data Archive, UKDA)数据生命周期管理模型指出^[5],科学数据生命周期包括数据管理、数据归档、数据格式化、数据存储、数据伦理验证5个阶段^[6],目前高校图书馆已开展的科学数据管理业务包含前4个阶段的服务,科学数据验证作为数据生命周期的最后一个阶段,高校图书馆也应参与其中并提供相应的验证服务。科学数据验证业务的主要功能是科研诚信的防范和补救,是科研不端行为的“事前预防”,以及在发现科研不端行为后及时采取相应的措施,尽量降低不端行为的危害和影响^[7]。高校图书馆拓展科学数据验证业务有利于充分利用图书馆的信息资源和充分发挥图书馆的信息服务功能,扩大图书馆的影响力,提高图书馆的地位。

2 高校图书馆参与科学数据验证的可行性

在信息资源数字化背景下,高校图书馆职能已从传统的对图书采集、编目、收藏向为用户提供个性化信息服务转变,高校图书馆的角色也由参与数据生命周期的个别阶段向服务内容贯穿整个数据生命周期延伸。

2.1 越来越多地参与科学数据管理

目前很多科学数据验证都依赖于同行评议,但现实是科学数据错误、被伪造、被剽窃等现象日益显著,其原因之一是我国没有进行科学数据审查验证的中介机构,完全依赖于同行评议系统。然而在实际操作中,同行评议者会比较关注数据量是否足以得出显著性结果,而不太会去检测是否是欺骗性结果^[8],一方面,同行评议者往往很难发现科学数据不诚信的问题,他们缺乏科学数据验证的数据资源、没有足够的时间去调查验证,有时同行评议者甚至难以辨别是否是原始数据;另一方面,来自不同学科领域的同行评议者在判断数据是否不诚信时容易观点相歧,也在某种程度上影响了科学数据审查与验证的效果。

近来,越来越多的高校图书馆建立了科学数据管理服务平台,提供科学数据管理服务,科学数据管理服务为高校图书馆拓展数据验证业务提供了良好的机遇和优势。图书馆的科学数据管理内容包括推行数据管理政策和标准、完善信息资源建设结构和体系、进行数据加工、融合、挖掘和分析、建立数据存储系统等^[9]。科学数据管理平台存有大量科研人员的原始数据为图书馆拓展科学数据验证业务提供了前提条件,弥补了现有的同行评议难以获得原始数据的缺陷,另外,图书馆拥有的多学科背景的专业数据管理人员,能够对数据进行定性定量分析、关联分析、数据挖掘等操作,易于发现数据错误

并进行核查,一定程度上保证了图书馆科学数据验证业务开展的科学性。

2.2 具备数据验证的业务基础

高校是大量科学数据的产生地,图书馆是高校科学研究活动的重要服务支撑机构,也是高校参与科学数据验证的最佳部门,目前高校图书馆的科技查新业务越来越受重视,科技查新业务作为科研活动生命周期的开端,已成为理工科科研立项必不可少的重要环节之一,而科学数据验证作为科研活动生命周期的末端阶段,也受到了越来越多的关注。高校图书馆已开展的科技查新业务和学科服务是其参与科学数据验证的重要业务基础。

国内外越来越多的高校图书馆开展科研诚信和数据管理的教育课程和培训。英国剑桥大学^[10]、牛津大学^[11]网站都有专门的关于“research integrity”的介绍,内容包括关于科学研究实践的详细标准、学校对于科研活动的监管政策和数据管理的培训等,也有关于科学数据验证的部分内容。学校的图书馆、行政办公室、其他机构以及全校师生均是科学数据的监察者和验证者,可以对科学数据进行监护、提出质疑并要求验证数据,每位研究人员必须严格遵守学术道德和学校政策;牛津大学开设了“Research Integrity Online”课程^[12],剑桥大学图书馆非常重视对研究人员科学数据验证的培训,为了提高考古学一年级博士研究生管理科学数据的技巧,培养科学研究的学术道德规范,开设了科学数据管理培训课程(DataTrain: open access post-graduate teaching materials in managing research data in archaeology)^[13],学生通过课程学习可以拿到相应的学分。国内武汉大学图书馆在对研究人员进行学术道德教育方面走在前列,武汉大学图书馆面向全校硕士研究生新生,开设了1学分的必修课程《研究生学术道德和学术规范》^[14],这些课程内容都涉及数据验证。

2.3 反哺科研产出

目前我国尚无对科学数据验证的专业机构,科学数据的发布或发表缺乏统一的规范程序,只需经过专家审查程序即可,审查方式过于单一,虽然学术论文中的科学数据仍需同行评议,但是弊端依然很多,如上文中提及的同行评议造假行为等,极大程度地影响了科研成果的产出、分享和后续阶段的转化利用效果。高校科学研究活动依托于图书馆的资源和服务,图书馆也可以为科研人员提供科学数据验证服务,利用图书馆的大量资源对科学数据的完整性和真实性进行验证,既让科研人员重新复查自己的科研成果,也有效规避数据伪造等学术不端行为。科研人员可以对有问题的科学数据及时修改

或删除,从而让高质量的科学数据得以保存、发表和转化利用。图书馆以此为基础创建高质量的科学数据库,反哺科研产出,可以让科研成果更具权威性和发挥更大的价值效用,提升科学研究的社会价值和经济效益,并进一步让我国学术成果走向国际,扩大我国学术的国际影响力。诚然,这也是高校图书馆主动承担部分学术不端责任的具体体现。

3 高校图书馆参与科学数据验证的途径

高校图书馆参与科学数据验证需要通过创建科学数据验证平台来实现,科学数据验证的具体途径应包括图书馆员角色的转变、提供科学数据验证服务和参与科学数据管理整个生命周期等。

3.1 从学科馆员到数据馆员

自1998年清华大学图书馆开始实行学科馆员制度,经过10余年的发展历程,学科化服务已在资源建设、队伍建设等诸多方面取得了长足的进步,2008年中国科学院文献情报中心初景利、张冬荣提出“第二代学科馆员”概念,提供全程服务内容包括课题策划、创新性论证、研究过程、论文发表、成果评价等^[15];张晓林提出“学科馆员3.0”的设想,是基于用户的、覆盖知识能力和嵌入科研过程的知识服务,强调的是用户需求服务^[16];武汉大学图书馆副馆长张洪元预测学科服务未来发展的最佳可能是:以文献为基础的科研分析、科研评价甚至科研指导将逐步变为现实,并成为高层次学科服务的主流内容^[16]。高校图书馆参与科学数据验证将对图书馆员能力提出更高的要求,图书馆员要做好从学科馆员到数据馆员的角色转变,在当前大数据科研环境背景下,数据能力是学科馆员的必备能力之一,数据验证将成为学科馆员的工作内容^[17]。

科学数据验证对象通常是针对某一专业领域前沿科学的数据,因此,数据馆员除了熟悉图书馆员基本业务内容外,必须具备相关专业的学科知识背景,了解学科发展动态和学科发展前沿,具备学科服务能力;科学数据验证必然要运用数据分析和验证的工具,对大量的原始数据进行检索、组织、校对和存储,因此,数据馆员要具备使用数据验证工具的能力和对科学数据进行管理的能力;科学数据验证目的是为了帮助科研人员查找到错误数据和避免数据造假、剽窃、抄袭、伪造等学术不端行为,这要求数据馆员要十分熟悉学术道德和学术伦理并能严格遵守学术研究规范,相信科学但敢于怀疑权威,在数据审核时做到公平、公正;最后,数据馆员必须具备良好的科学数据素养,广义的科学数据素养包括对科学研究活动中数据的收集、描述、组

织、管理、评价和利用数据的知识和能力,强调对科学数据产生、操作和评价的能力^[18]。高校图书馆参与科学数据验证要求图书馆员主动做好角色的转变,不断加强自身业务能力和思想道德建设,自觉遵守学术道德伦理和学术规范,不断提高自身科学数据素养,为图书馆拓展科学数据验证业务做好准备。

3.2 提供科学数据验证服务

高校图书馆开展一项崭新的业务前必须做好业务规划,科学定位,明确目标,优化内部技术组合。图书馆开展数据验证业务:首先,需要政策法规的支持,只有用政策法规来确保图书馆的地位和业务的权威性,验证结果才能被科研人员信服;其次,图书馆要加强人力资源建设,做好人才贮备,招聘相关专业人员,一方面,开展数据验证服务需要图书馆各部门的通力协作,要提高图书馆各部门馆员的信息服务能力,另一方面,数据验证馆员的服务能力直接影响了数据验证服务的进程和效果,这也对图书馆人才贮备提出了更高的要求;最后,图书馆为适应科学数据验证业务的新需求,应当提前做好馆员的专业课程教育和业务培训^[19],快速提高馆员的科学数据验证业务能力。另外,图书馆科学数据管理服务尚处于起步阶段,图书馆从传统的文献服务到科学数据管理服务、科学数据验证服务仍然有很大的发展空间^[20]。图书馆要继续加大现有的业务优势,对图书馆馆藏资源、网络资源和共享资源进行整合与揭示,优化馆藏结构,加强图书馆内部各部门的协调合作,引进新技术(如虚拟现实技术、无人机技术),不断优化图书馆技术组合,为研究人员提供科学数据验证服务。

3.3 参与科学数据管理的整个生命周期

高校图书馆为科学数据管理生命周期的不同阶段提供不同层次的服务内容,主要可分为传统业务层(包括参考咨询、科技查新业务等)、正在培养的业务层(包括学科化服务、科学数据管理服务)、待开发业务层(包括科学数据验证业务等)、学术环境层(包括学术不端检测、净化学术环境)和政策管理层(净化学术环境、科研政策导向)。高校图书馆参与科学数据管理生命周期流程见图1。

高校是大量科学数据的产生地,参考咨询、科技查新业务等是高校图书馆的基础业务内容,是高校图书馆参与科学数据管理的开端,参考咨询是科研活动的重要辅助手段之一,科技查新业务可以检测篡改数据、捏造事实等学术不端行为;近来越来越多的高校图书馆实行学科服务制度,学科服务是