

世界主要国家科学数据资源共享和管理的 对比分析和启示

杨行¹ 屈宝强² 赫运涛³ 赵伟²

(1. 首都师范大学图书馆, 北京 100089; 2. 中国科学技术信息研究所, 北京 100038;
3. 国家科技基础条件平台中心, 北京 100862)

摘要: 科学数据是信息时代重要的战略资源, 世界各国对科学数据资源的重视程度也越来越高。通过网络调研、链接分析以及案例分析等方法, 从科学数据资源的规模、科学数据中心的国际参与度、科学数据共享平台的网络显示度以及科学数据出版的成熟度等4个方面综合分析世界主要国家在科学数据共享领域的发展现状, 并结合中美两国地球科学领域科学数据共享平台的案例分析, 发现我国在该领域与世界先进水平的差距, 为提升我国科技资源管理与共享能力提供决策支持。

关键词: 科学数据; 研究数据; 数据出版; 资源共享

中图分类号: G203

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2016.06.003

Contrast Analysis and Its Revelation from Scientific Data Resource Sharing and Management on Main Countries

YANG Xing¹, QU Baoqiang², HE Yuntao³, ZHAO Wei²

(1. Capital Normal University, Beijing 100089, 2. Institute of Scientific and Technical Information of China, Beijing 100038; 3. National Science and Technology Infrastructure Center, Beijing 100862)

Abstract: Scientific data plays an important role in the information era, nations around the world is putting more and more emphasis on scientific data resources. Network survey, link analysis and case study are applied to make a quantitative analysis in order to study the development status of scientific data resource sharing around the world's key countries. This survey contains the following four aspects: scale of scientific data resource, data centers' participation in related international organizations, web visibility of scientific data sharing platforms, and maturity of scientific data publishing. For the purpose of providing decision support for our scientific data management and sharing, this paper also does a case study between Chinese and American Geodata sharing platforms.

Keywords: scientific data, research data, data publishing, resource sharing

作者简介: 杨行* (1990—), 女, 首都师范大学图书馆助理馆员, 硕士研究生, 研究方向: 信息资源管理与共享; 屈宝强 (1980—), 男, 中国科学技术信息研究所研究员, 博士, 研究方向: 信息资源管理; 赫运涛 (1980—), 男, 国家科技基础条件平台中心副研究员, 研究方向: 科技资源管理与共享; 赵伟 (1975—), 女, 中国科学技术信息研究所研究员, 博士, 研究方向: 科技资源共享与评价。

基金项目: 国家社会科学基金一般项目“科学数据引用的实现机制研究”(15BTQ053); 国家科技基础条件平台专项课题“科技基础条件资源指数国际比较分析研究”(2016DBJ1ZZ04)。

收稿时间: 2016年9月7日。

1 引言

科学数据资源是信息时代最基本、最活跃的科技资源，对科技进步与创新、经济增长、社会发展以及国家安全发挥着关键作用。广义的科学数据资源既包括在各类科技活动中产生的原始性、基础性数据，也包括各部门通过大规模观测、探测、调查和试验工作所获得的海量数据，甚至包括科学家在工作中长期积累的个人数据^[1]。无论是业务型科学数据还是研究型科学数据，都是科学数据共享过程中必须进行筛选、整合、集成，并为科技创新提供支持服务的数据资源^[2]。随着科学数据保藏与共享理念的普及以及大数据技术的不断发展，科学数据资源的科学价值和使用价值得以充分体现，科学数据资源的建设、管理与共享工作也逐渐得到了世界各国政府、科研机构和科学家的高度关注和重视。世界范围内面向科学数据管理和共享的专门机构相继成立、各国科学数据共享领域参与主体（如国家科学数据组织、政府、科研资助机构、学术出版机构等）对科学数据共享项目的资金支持力度逐渐加大，全球学术出版团体对科学数据共享政策的要求也日益规范，并积极开展对科学数据出版实践的探索工作。

目前，学术界对国际科学数据管理与共享发展的研究多以定性分析为主^[3-4]，定量研究较少。为准确把握科学数据共享所带来的机遇和挑战，有必要持续追踪并研究国际科学数据共享领域的新进展，从数据维度客观反映世界主要国家在科学数据共享领域的发展差异。

2 数据来源与方法

本文采用网络调研、链接分析以及案例分析等方法，对科学数据领域重要国际组织（如世界数据系统 WDS、DataCite）、科学数据中心以及数据期刊官方网站等公开渠道发布的数据进行汇总整理，并进行定量分析。

调查内容主要包括科学数据资源规模、科学数据中心、科学数据存储与共享平台以及科学数据出版等 4 个方面，重点涉及美国、加拿大、日

本、德国、法国、英国、意大利、澳大利亚等发达国家，以及中国、俄罗斯、印度、巴西、南非和韩国等新兴经济体，共计 14 个国家。

3 主要国家科学数据资源共享现状分析

3.1 科学数据资源的建设规模

各学科领域内产生的科学数据资源量虽然非常庞大，但明显存在质量参差不齐的问题。同时，科学数据资源因其自身的特殊性，分散在不同国家地区、不同研究项目以及不同科学家手中，对其绝对数量的调查存在较大困难。为尽可能全面、准确地考察世界主要国家科学数据资源的建设规模，利用 DataCite 以及中文 DOI 的官方数据，借助各国科学数据的 DOI 注册量来反映各国优质科学数据资源的建设规模。

图 1 为世界主要国家科学数据资源的 DOI 注册情况对比。除中国的数据来自中文 DOI 外，美、德、英等国家科学数据资源的 DOI 注册总量来自于 DataCite 的统计数据，在调查对象范围内的俄罗斯、印度、巴西、南非和韩国等国暂未获取到相关数据。不难看出，美国、德国科学数据资源的 DOI 注册总量遥遥领先，这与两国长期以来对科学数据资源建设与管理的高度重视密切相关；英国的 DOI 注册数量虽仅排第三，但较加拿大、意大利、澳大利亚等国仍有较大优势。我国科学数据资源的 DOI 注册数量仅高于日本，在世界主要国家中排第八位，仍有较大提升空间。

3.2 科学数据中心的国际参与度

科学数据中心是科学数据管理与共享领域内实施具体事务、解决实际问题、促进科学数据资源管理与共享水平持续提高的专门机构。科学数据中心在国际范围内的广泛参与有助于其资源在更宽广的领域、地域内获得更高效的共享与利用。此次调查主要通过各国科学数据中心加入国际组织的情况来考察各国科学数据中心的国际参与度，重点考察世界主要国家在世界数据系统（WDS）的会员单位数量以及在 DataCite 注册过的科学数据中心数量。其中，世界数据系统（<https://www.icsu-wds.org>）是于 2008 年由全世

界范围内的数据中心、数据服务组织共同组成,目前已经覆盖了全球范围内的90多家数据中心或数据服务组织,涉及美国、法国、德国、中国、荷兰、俄罗斯、英国、澳大利亚等20多个国家和地区,涵盖地圈、生物圈、外层空间、天文现象、生物多样性、生态环境、人文、健康、遥感和土地覆被等多个学科领域。而DataCite作为一个将研究人员、学者、数据中心、图书馆、出版商以及基金赞助者等利益相关方汇聚起来以支持可靠、永久数据识别的会员组织,其成员同样分布在全球各地,并通过其成员机构在世界范围内提供服务。

如图2所示,美、德两国加入科学数据领域国际组织的数量较其他国家依然有绝对优势。统计结果还显示,世界数据系统的成员单位中近1/3来自美国。英国、澳大利亚、法国、意大利、加拿大等5个国家科学数据中心的发展则基本处于第二阵营,第三阵营则主要以中国、俄罗斯为首。目前,我国加入世界数据系统的科学数据中

心共6个,分别是:天文数据中心、可再生资源与环境学科中心、世界海洋数据信息服务、世界微生物数据中心、空间科学数据中心以及寒区旱区科学数据中心。

值得一提的是,世界主要国家科学数据中心加入国际组织的情况与各国科学数据资源的DOI注册总量分布基本一致。可见,科学数据中心的发展壮大与科学数据资源的持续建设之间必然存在良性的因果循环。

3.3 科学数据共享平台的网络显示度

科学数据共享平台是实现科学数据存储管理、促进科学数据开放获取与重复利用、支撑科学研究的关键基础设施^[5]。近年来,随着国际科学数据组织的推进与各国政策法律的支持,国内外科学数据共享平台的建设工作不断加强。Re3data.org是由德国研究基金会(DFG)资助建设的全球性研究数据仓储注册系统。截至2016年9月1日,在该系统内注册的全球科学数据仓储/共享平台共1688个,世界主要国家科学数据

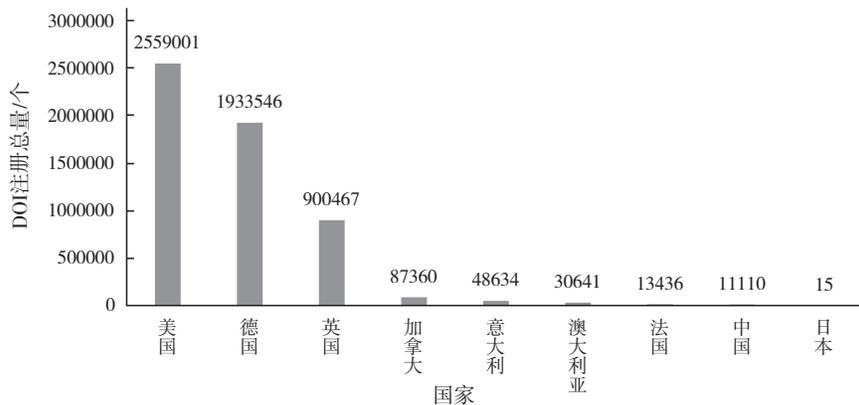


图1 世界主要国家科学数据资源的DOI注册情况对比

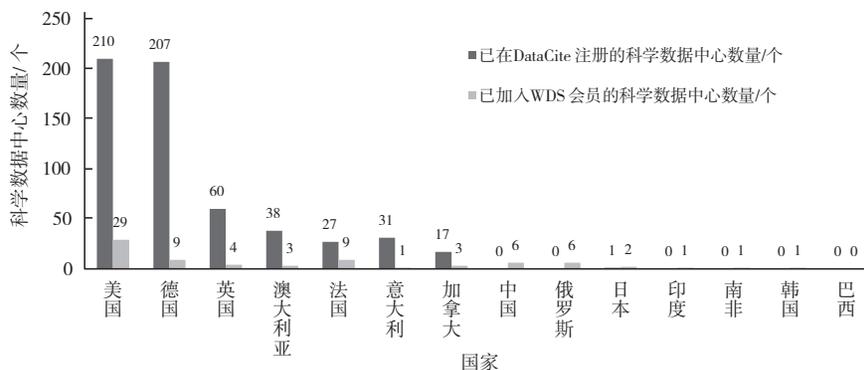


图2 世界主要国家科学数据中心加入WDS及DataCite的情况对比

共享平台的建设情况如图 3 所示。

各国科学数据共享平台的建设数量在一定程度上能够反映各国对科学数据管理工作的重视程度。为进一步考察各国科学数据共享平台建设的实际情况，采用链接分析的方法，从各国众多科学数据共享平台中随机选取 5~10 个样本，利用平均外部链接数指标，考察世界主要国家科学数据共享平台在网络范围内的显示度^[6]。必须说明的是，虽然平台建设质量越高，其网络显示度未必越高，但一般可以认为，网络显示度较高的平台应基本处于正常运行的状态。

图 4 是各国科学数据中心平均外部链接数的对比，可以看出我国科学数据共享平台在网络范围内的显示度较低，与其他国家有较大差距；各国平均外部链接数为 14201 个，我国科学数据共享平台的网络显示度与该平均水平仍有较大差距；与链接数最多的德国相比，相差两个数量级。为兼顾公平公正的原则，此次调查在评估各

国科学数据共享平台时测试的均是各参评平台的英文版本，受语言限制，中国、日本、韩国、俄罗斯等国的评估结果可能会略逊于实际运行情况。但也从另一方面表明，我国科学数据资源事业在未来若要走向国际舞台，除建设好本土共享平台外，还应当兼顾英文平台的建设与推广。

3.4 科学数据出版工作的成熟度

科学数据出版能够使数据达到可引用和追溯的状态，其核心内容是为数据引用提供标准的数据引用格式和永久访问地址^[7]，是科学数据开放共享的一种创新模式。科学数据出版有望实现全面数据质量管理，能有效保持数据的完整性，防止科学造假，还能减少重复时间及资金投入、提高数据发表者或发表机构在科研领域的知名度和影响力，在减少跨学科研究障碍以及提高科研合作几率等方面均具有重大意义。

数据出版主要包括 3 种形式：（1）数据论文（Data Paper），重点描述科学数据本身，不涉

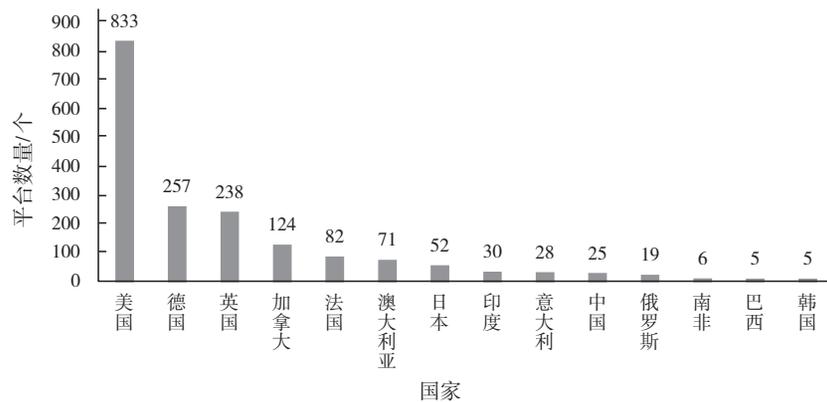


图 3 世界主要国家科学数据共享平台的建设情况对比



图 4 世界主要国家科学数据共享平台的平均链接数对比

表1 世界范围内代表性数据期刊

序号	期刊	ISSN	出版商	国家	时间范围
1	Earth System Science Data	1866-3508	Copernicus Publications	德国	2009-2015
2	Scientific Data	2052-4436	Nature Publication Group	英国	2014-2015
3	Geoscience Data Journal	2049-6060	Wiley Online Library	美国	2014-2015
4	Biodiversity Data Journal	1314-2828	Pensoft Publishers	保加利亚	2013-2015
5	Genomics Data	2213-5960	Elsevier	荷兰	2014-2016
6	Dataset Papers in Science	2314-8497	Hindawi Publishing	埃及	2013-2015
7	International Journal of Robotics Research	1741-3176	SAGE Publications	美国	2009-2015
8	F1000 Research	2046-1402	F1000 Research	英国	2012-2015
9	GigaScience	2047-217X	BoiMed Central	英国	2012-2015
10	SpringerPlus	2193-1801	Springer	美国	2014-2015
11	Journal of Open Archaeology Data	2049-1565	Ubiquity Press	英国	2012-2015
12	Journal of Open Psychology Data	2050-9863	Ubiquity Press	英国	2013-2015
13	Ecology	0012-9658	Ecological Society of America	美国	1999-2015

及科学假设以及科学问题的研究结果,是接受同行评议的、正式出版的学术论文。目前,国内外发表数据论文的期刊主要包括混合型(同时出版普通论文和数据论文)、单一型(仅出版数据论文)两种。(2)独立的数据出版,指数据中心对本机构保藏数据的独立出版,如各国政府数据中心(data.gov)。(3)论文辅助资料,主要以附件形式存在,或将数据存储在专业数据库或出版机构网站上,作为支撑论文学术思想的佐证。

此次调查主要针对单一型数据期刊出版进行,利用各国在国际数据期刊上发表的数据论文数量来反映各国在科学数据出版领域的成熟度,主要涉及地球科学、生命科学等领域,参与调研的13种数据期刊基本信息如表1所示。

图5是世界主要国家科学数据论文的出版情况对比。截至2015年年底,中国学者在上述国际重要数据期刊发表的数据论文共计66篇,仅次于美国(265篇)、英国(107篇)和德国(88篇),且这一数据还未包括中国学者在国内数据期刊上的发布数量。可见,国内学者已经逐渐意识到数据共享的积极作用,并愿意将研究数据发布出来。国内数据出版期刊的建设工作仍处在摸

索阶段,虽然我们并未将国内数据期刊纳入统计范畴,但我国的数据期刊建设工作已初见成效。例如:自2011年以来,《生物多样性》开设了生物编目栏目,刊发生物多样性热点地区、调查空白地区的物种编目以及中国和世界重要生物类群的编目;《植物生态学报》开辟了资料论文专栏,旨在刊发特定尺度上(如群落、景观或区域)植物群落的样方数据;《地理学报》从2014年开始以增刊的形式刊发数据论文,并将数据汇交至数据平台;致力于开展多学科领域科学数据出版的《中国科学数据》于2015年已获得国家新闻出版总署批准发行。

4 美国地球科学领域数据共享平台案例分析

国家地球科学数据共享平台和地球观测系统数据信息系统分别是中美两国地球科学领域非常具有代表性的数据共享平台^①。本文采用案例分析的方法,从微观层面对比两国在数据共享领域工作成效方面的差异。

4.1 建设目标及基本架构

美国地球观测系统数据信息系统(Earth Observing System Data and Information System,

^① 美国地球观测系统数据信息系统的访问地址为<https://earthdata.nasa.gov/>;中国国家地球科学数据共享平台的访问地址为<http://www.geodata.cn/>。

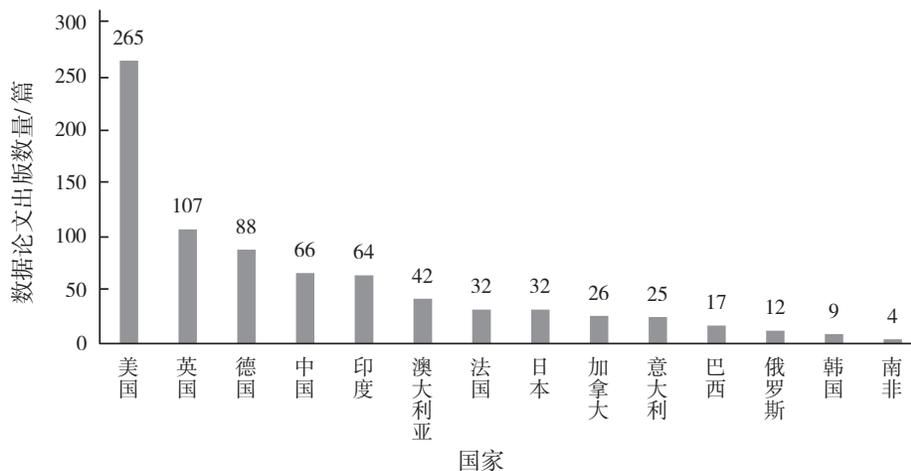


图5 世界主要国家科学数据论文的出版情况对比



图6 美国地球观测系统数据信息系统资源量分布

EOSDIS)属于EOS地面系统范畴,主要负责地球科学数据存档、产品制作和分发。该系统共有12个下属的分布式数据存档中心DAAC(Distributed Active Achieved Center),分别负责制作地球观测系统观测得到的各类、各级数据产品,并进行存档管理,为不同研究领域的用户提供各具特色的数据产品、数据信息和数据使用工具等方面的服务^[8]。各数据存档中心分布于不同地点,但所有数据通过EOS元数据信息交换站(EOS Clearing House, ECHO)进行统一管理。

中国国家地球系统科学数据共享平台的总体

目标是以元数据为核心整合集成分布在国内外数据中心群、高等院校、科研院所以及科学家个人产生的数据资源,引进国际数据资源,接收国家重点科研项目产生的数据资源,并在此基础上生产加工数据产品。该平台按照“总中心—数据中心—数据资源点”三级架构组织实施,形成了由1个总中心、6个学科数据中心和8个区域数据中心构成的物理上分布、逻辑上统一的一站式数据共享服务网络系统。其数据资源涉及地球系统的各个圈层、领域、学科和特色区域^[9]。

① 系统使用情况相关数据来自于EOSDIS官方网站,各领域资源量数据根据官方网站检索数据统计得出。

② 各领域资源量数据根据国家地球系统科学数据共享平台官方网站检索数据统计得出。

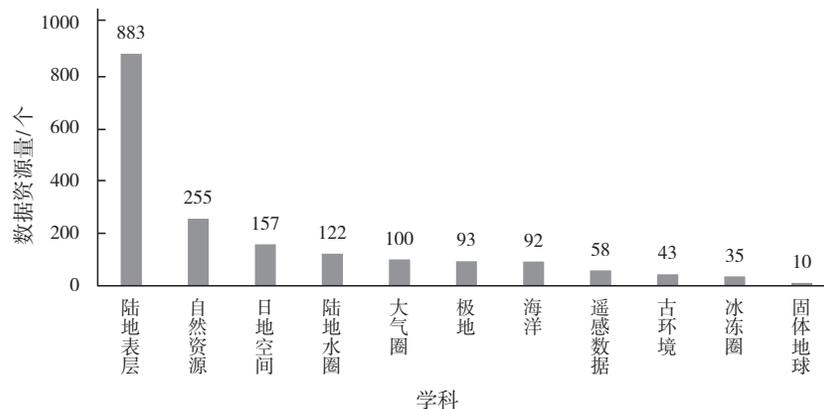


图7 (中国) 国家地球系统科学数据共享平台资源量分布

4.2 资源建设及使用情况

2015 财政年期间 (2014 年 10 月 1 日—2015 年 9 月 30 日), 美国地球观测系统数据信息系统共新增数据集 9462 个, 网站页面访问量共计 240 万人次, 数据资源总量为 14.6PB, 日均增长量达 16.0TB, 年内共计向 14.2 亿个终端用户提供数据产品, 平均每天向终端用户提供的数据量达 32.1TB。该系统各领域内的数据资源量分布如图 6 所示^①。

自正式上线至 2014 年年底, 我国国家地球系统科学数据共享平台实名注册用户共计 9.4 万多名, 网站总访问人次 1800 多万; 向科技界和社会公众提供了 93.53TB 的数据服务量。为 2384 项重大科研项目/课题, 包括国家 973 项目、国家科技支撑计划、国家自然科学基金等各类国家和省部级科研项目, 35 项重大建设工程, 34 项民生工程提供了有效的数据服务。该平台目前共有数据资源 59.56TB, 涉及大气圈、陆地表层、陆地水圈、冰冻圈等 11 个一级类, 各学科领域内的数据资源量分布如图 7 所示^②。

我国自 2001 年有组织地开展地球科学领域的科学数据共享工作, 虽然已经取得了较为丰硕的成果, 但相比起步较早的美国仍存在较大差距。美国地球科学观测系统数据系统不仅在资源拥有量上远超中国, 在数据服务方式上也相对成熟。除提供全球范围内地球科学数据的浏览、检

索、下载服务外, EOSDIS 还为用户提供各类技术支持与指导, 帮助用户选择并获取数据, 提供数据处理及可视化工具, 并及时告知用户与数据相关的更新信息。而我国国家地球系统科学数据共享平台目前的工作重点仍是整合共享陆地表层系统和人地关系研究所需要的数据资源, 并重点开展不同层面的数据揭示工作, 在对数据资源的描述、引用、关联等方面的工作也日趋成熟。

5 结语

总体而言, 我国是一个科学数据大国, 但还不是一个科学数据强国, 在科学数据管理与共享领域方面的工作起步较晚, 目前该领域的发展状况与美、德、英等发达国家仍有较大差距。然而, 自 2002 年正式启动“国家科学数据共享工程”项目以来, 我国在科学数据共享与管理领域已取得了长足进步, 各项工作也逐渐与国际接轨。不仅开始重视对数据资源唯一标识及永久访问方面的工作, 而且积极支持科学数据中心以及科学数据共享平台的建设。除资金方面的大力支持外, 还制定了一系列科学数据共享相关的政策和标准, 在科学数据出版方面的实践成果也已崭露头角。

科学数据资源共享的本质是满足广大用户对稀缺数据资源的需求^[10]。科学数据共享事业的健康发展离不开用户的支持。因而, 科学数据基础设施, 尤其是科学数据共享平台应当基于科学数

据用户的需求和体验不断优化资源组织方式、创新数据服务模式。

参考文献

- [1] 傅小锋,李俊,黎建辉,等.国际科学数据的发展与共享[J].中国基础科学,2007,9(2):30-35.
- [2] 王卷乐,祝学衍,石蕾,等.国际研究数据联盟及对我国科学数据共享的启示[J].中国科技资源导刊,2014(2):15-20.
- [3] 李娟,刘德洪,江洪.国际科学数据共享现状研究[J].图书馆建设,2009(2):25-27,31.
- [4] 邱春艳,黄如花.近3年国际科学数据共享领域新进展[J].图书情报工作,2016(3):6-14.
- [5] 赵伟,彭洁,杨行,等.我国科学数据共享网站评价研究[J].中国科技资源导刊,2014(2):1-6.
- [6] 邱均平,马凤,曾倩.中国大学的网络产出、显示度和影响力研究[J].情报科学,2011(3):325-332.
- [7] 吴立宗,王亮绪,南卓铜,等.科学数据出版现状及其体系框架[J].遥感技术与应用,2013(3):383-390.
- [8] 王旻燕,臧海佳,邓莉.NASA地球科学数据分布式数据存档中心的数据和数据管理[J].气象科技合作动态,2009(1):1-9.
- [9] 王卷乐,宋佳,卜坤,等.国家地球系统科学数据共享平台数据分类编目与特征分析[J].中国科技资源导刊,2015,47(6):65-73.
- [10] 彭洁,赵伟,屈宝强.基于用户视角的科技资源开放共享评价理论模型研究[J].中国科技资源导刊,2013(2):1-5.

中国科技信息事业创立60周年座谈会在京召开

本刊讯 庆祝中国科技信息事业创立暨中国科学技术信息研究所创建60周年座谈会于2016年10月15日在北京举行。全国政协副主席、科技部部长、中国科协主席万钢,科技部党组书记、副部长王志刚,科技部党组成员、副部长李萌,党组成员、科技日报社社长李平出席座谈会。国家各部委、省市情报机构、高等院校和科研院所专家代表等110余人参加了座谈会。

王志刚书记在讲话中指出,60年来,我国科技信息事业为国家科学技术发展、改革创新实践发挥着“助推器”和“加速器”的作用,为推动科技、经济与社会的发展以及国家安全作出了重要贡献。面对国家赋予科技创新伟大而重要的使命,科技信息要成为贯彻实施好创新驱动发展战略的先行官。

在新的历史阶段,贯彻实施好创新驱动发展战略,要坚持面向世界科技前沿、面向国家重大需求、面向国民经济主战场,抓紧完成构筑先发优势、推动大众创业万众创新、强化科技基础、拓展创新空间、完善创新治理体系、营造良好创新生态六大战略部署,落实好科技重大项目和工程、前沿探索、原始创新、保障科技供给、深化科技体制改革、弘扬创新精神等重大任务。

王志刚书记强调,创新驱动,情报先行。新时期加强科技信息工作,既是支撑科技创新的重要基础性工作,更是科技信息服务机构主动适应行业发展新特

点、新趋势、新要求的必然之举。国家正着力完善科技创新基础制度,建立公开统一的国家科技管理平台,加快建立健全国家科技管理信息系统,科技信息行业要承担起这一重要的历史使命,围绕全面参与《国家大数据发展行动》,共同建设国家政府数据统一开放平台,成为引领大数据资源建设的战略制高点。要围绕全面支撑创新治理和“双创”,开展面向创新全链条的科技情报与信息服务,成为推动科技服务业发展的主要贡献点。要围绕全面落实《深化科技体制改革实施方案》,深化科技信息服务机构和行业改革,成为展示科技改革成效的重要支撑点。要围绕全面营造优良的创新创业环境,积极认真履行公益性研究机构的社会责任,成为提升全社会公民科学素养的关键承载点。

中国科学技术信息研究所是在党和国家领导人亲切关怀和指导下成立并发展起来的新中国第一批国家级科研院所,开创了我国的科技信息事业,持续引领全国科技情报事业的发展。60年来,该所科学研究、公益服务、决策支持和技术研发能力持续提升,先后9次获国家科技进步奖,其中一等奖两次。迈入新时期,中国科学技术信息研究所将携手国内外同行,以更加开放的理念,创新的精神,为各类创新主体提供优质高效的全方位科技信息与决策支持服务,向着“国家科技创新高端智库、国家科技信息大数据中心、国家科技信息资源综合利用与服务中心”的目标前进。