

湖北省学术著作出版专项资金资助项目

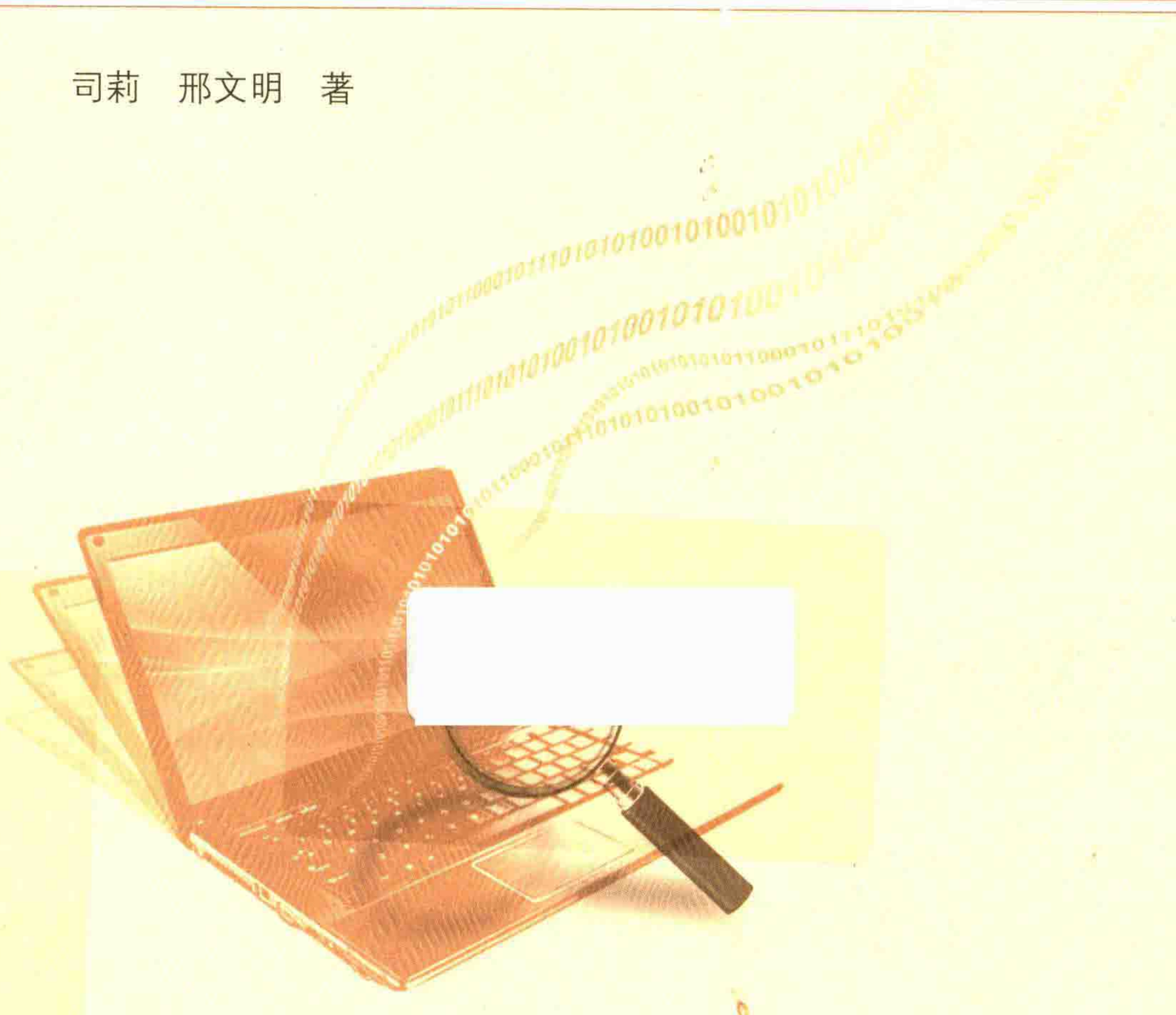


数字信息资源与服务创新研究丛书

科学数据管理与共享的 理论与实践

Theory and Practice of
Scientific Data Management and Sharing

司莉 邢文明 著



WUHAN UNIVERSITY PRESS

武汉大学出版社

湖北省学术著作出版专项资金资助项目

 数字信息资源与服务创新研究丛书

科学数据管理与共享的 理论与实践

司莉 邢文明 著



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

科学数据管理与共享的理论与实践/司莉等著. —武汉: 武汉大学出版社, 2017.3

数字信息资源与服务创新研究丛书

ISBN 978-7-307-17933-2

I.科… II.司… III.科学研究—数据管理—研究 IV.G202

中国版本图书馆 CIP 数据核字(2016)第 122296 号

责任编辑:詹 蜜

责任校对:李孟潇

版式设计:马 佳

出版发行: **武汉大学出版社** (430072 武昌 珞珈山)

(电子邮件: cbs22@whu.edu.cn 网址: www.wdp.com.cn)

印刷: 武汉精一佳印刷有限公司

开本: 720×1000 1/16 印张: 18.75 字数: 269 千字 插页: 8

版次: 2017 年 3 月第 1 版 2017 年 3 月第 1 次印刷

ISBN 978-7-307-17933-2 定价: 59.00 元

版权所有, 不得翻印; 凡购我社的图书, 如有质量问题, 请与当地图书销售部门联系调换。

前 言

科学数据不仅是技术预见、科技创新、经济发展和国家安全的重要战略资源,也是政府部门制定政策、进行科学决策的重要依据。科学数据是信息时代科技创新的驱动力,新的科学机遇来自于越来越有效的数据组织、共享与利用。科学数据管理与共享已成为国家发展战略目标之一,受到国际组织、政府部门及研究机构的高度重视。联合国教科文组织(UNESCO)推出了“促进发展中国家科学数据共享与应用全球联盟”计划,国际科学理事会(ICSU)支持建立了两个鼓励促进全球科学数据共享的国际组织——国际科技数据委员会(CODATA)和世界数据系统(WDS),开展一系列的国际合作行动。美国最重要的科研资助机构国家科学基金会(NSF)要求项目申请书中须包含“数据管理计划”,美国研究图书馆协会(ARL)将科学数据管理与共享服务作为战略规划之一。英国研究委员会(RCUK)及其下属的7个研究理事会都制定了数据共享的相关政策,建立了数据仓库(中心)供研究者上传其科研数据。同时,英国还成立了专门的“数据存储服务中心”(DCC)。科技创新是提高社会生产力和综合国力的战略支撑,我国已将之放在国家发展全局的核心

位置。党的“十八大”报告明确提出实施创新驱动发展战略。《国民经济和社会发展规划“十二五”规划纲要》提出要加强基础信息资源体系,强化信息资源的整合。2012年7月温家宝在全国科技创新大会上指出要建立科技资源开放共享机制,国家支持的科研活动所获得的信息资料,要最大限度地向社会公开。成立于2012年12月8日的全国科学院联盟签署《全国科学院联盟成立北京宣言》,提出要进一步集聚优质创新资源,共建协同创新体系,实现创新驱动发展。当前的e-Science环境与大数据时代带来了科学研究的信息化,科学研究已成为数据密集型活动,科技创新将越来越依赖于科学数据支撑。科学研究需要支持过程研究和综合研究以及在时间与空间上成序列、多要素融合的科研数据,如何保障科学数据的集成与共享,是实现创新驱动战略的首要课题。

科学数据管理成为一个新领域,在学术界和各行业发展中显现出越来越重要的作用。随着数据密集型、数据驱动的科研成为新的科研方式,科学数据的管理和共享迅速得到了国内外图书情报界、信息科学和其他学科领域的高度重视。我们采用文献调研法、实地考察与访谈法、实证分析与典型案例法和专家研讨与德尔菲法等研究方法,围绕科学数据管理中的研究与实践进展、需求、政策、数据组织与元数据标准、图书馆科学数据服务、数据管理与服务的绩效评估,以及科学数据人才培育等方面对科学数据的管理与共享进行了较深入的探讨。《科学数据管理与共享的理论与实践》是教育部人文社会科学研究项目(项目批准号:13YJA870016)研究成果之一。本书主要包括以下内容:①科学数据管理与共享:国际组织动向;②国内外科学数据管理与共享研究进展;③我国科学数据管理与共享的需求分析;④科学数据管理与共享的政策研究;⑤科学数据的组织与元数据标准;⑥图书馆开展的科学数据服务;⑦科学数据管理与共享的绩效评估;⑧科学数据管理的人才培育。

本书共八章。全书由司莉组织编写、统稿、审稿并定稿。撰写者为司莉、邢文明、黄如花、庄晓喆、邱春艳、李鑫、李月婷、辛娟娟、华小琴、吴方枝、王思敏、钱琦琪、郭卫宁。具体分工如下:第一章由黄如花、邢文明撰写;第二章由司莉、庄晓喆、王思敏、吴方枝、钱琦琪与李

鑫撰写;第三章由邢文明撰写;第四章由司莉、邢文明与李月婷撰写;第五章由邱春艳、司莉与李鑫撰写;第六章由邢文明、吴方枝与司莉撰写;第七章由司莉、李月婷、邢文明、华小琴、李鑫与辛娟娟撰写;第八章由司莉、邢文明、庄晓喆与郭卫宁撰写。

本书在编写过程中得到教育部人文社会科学项目资助,得到武汉大学信息管理学院与武汉大学出版社领导的大力支持,责任编辑詹蜜对于本书的出版付出了辛勤劳动,在此特表示诚挚的谢意。

限于专业视野与学术水平,书中难免会存在错误与不足,恳请专家、同行和读者批评与指正。

司 莉

2015年8月于武汉大学

C O N T E N T S 目 录

第 1 章 科学数据管理与共享:国际组织动向	1
1.1 国际组织关于科学数据的实践、会议与政策	1
1.1.1 国际组织关于科学数据共享的行动与项目	1
1.1.2 国际组织有关科学数据共享的会议	3
1.1.3 国际组织有关科学数据共享的政策与声明	5
1.2 国际组织在科学数据管理与共享中的作用	10
1.2.1 联合国是科学数据管理活动积极的推动者	10
1.2.2 欧盟的科学数据管理实践具有示范作用	13
1.2.3 经济合作与发展组织积极促进科学数据的开放和共享	16
1.2.4 国际科学技术数据委员会关注科学技术领域数据 管理与共享	17
第 2 章 国内外科学数据管理与共享研究进展	20
2.1 国外科学数据管理与共享研究进展	20
2.1.1 不同领域科学数据管理研究	21
2.1.2 科学数据管理方法与工具研究	22
2.1.3 关于科学数据共享的研究	27
2.1.4 关于图书馆开展科学数据管理服务的研究	31

2.1.5	关于科学数据管理人员教育与培训的研究	34
2.2	我国科学数据领域研究的主题分析	35
2.2.1	研究设计	36
2.2.2	研究的主要发现	37
2.2.3	讨论	43
第3章	我国科学数据管理与共享的需求分析	51
3.1	调查目的与方法	51
3.1.1	调查目的	51
3.1.2	调查对象与调查方法	52
3.2	调查问卷设计	52
3.3	调查结果	53
3.3.1	问卷回收情况	53
3.3.2	调查对象基本情况	53
3.3.3	调查问卷数据分析	56
3.4	调查结论	81
第4章	科学数据管理与共享的政策研究	85
4.1	国外科学数据管理与共享政策调查	85
4.1.1	科研资助管理机构制定的科学数据管理政策	85
4.1.2	高校制定的数据公开计划	90
4.1.3	政府部门制定的数据公开政策	103
4.2	国外科学数据保存政策的分析	107
4.2.1	研究方法	108
4.2.2	资助机构的科学数据保存政策内容分析	109
4.2.3	高校科学数据保存政策内容分析	112
4.3	国外科学数据管理政策对我国的启示	116
4.3.1	科研资助管理机构应出台相关规定	116
4.3.2	各高校应制定本机构的数据管理政策	116
4.3.3	政府应加大对科学数据共享的引导和支持	117
4.3.4	明确参与科学数据保存各方的权利与责任	117

4.3.5	保护研究成果,建立完整的权益保护体系·····	118
4.3.6	以节约科研资源、提高科学数据利用率为出发点·····	118
4.4	我国科学数据管理与共享政策的制定·····	119
4.4.1	政策制定的指导思想·····	119
4.4.2	政策制定的原则与目标·····	120
4.4.3	政策体系的构建·····	122
第5章	科学数据的组织与元数据标准 ·····	142
5.1	科学数据的元数据描述框架·····	142
5.1.1	科学元数据的研究现状·····	143
5.1.2	科学元数据的描述框架·····	149
5.2	图书馆参与科学数据管理中的元数据应用实践·····	158
5.2.1	科学数据管理相关机构对元数据的重视·····	159
5.2.2	图书馆科学数据管理基础服务对元数据的运用·····	161
5.2.3	自建科学数据平台对元数据的运用·····	164
5.2.4	科学数据管理合作中的元数据应用·····	167
5.3	英美政府数据门户网站科学数据组织与查询研究·····	169
5.3.1	英美政府数据门户网站数据组织、浏览与检索功能·····	170
5.3.2	英美政府数据门户网站数据组织和浏览、检索的特点·····	174
5.4	基于关联数据的科学数据集成与共享研究·····	176
5.4.1	国内外基于关联数据的科学数据管理研究现状·····	177
5.4.2	Bio2RDF 项目案例分析·····	181
5.5	期刊文献与科学数据的关联服务研究·····	186
5.5.1	关联服务的提供途径·····	187
5.5.2	关联服务的关键性实现方式·····	189
5.5.3	关联服务中的资源描述与组织·····	193
第6章	图书馆开展的科学数据服务 ·····	195
6.1	调查对象与方法·····	197
6.2	调查结果分析·····	197
6.2.1	调查结果·····	197

6.2.2	调查结果讨论与分析	206
6.3	对国内高校图书馆开展这一服务的启示	207
6.3.1	加强与利益相关者的合作是科学数据服务成功的关键	207
6.3.2	重视科学数据管理专门人才的教育和培养	209
6.3.3	充分利用相关资源与服务	210
6.3.4	重视数据引用和知识产权教育	210
第7章	科学数据管理与共享的绩效评估	212
7.1	我国科学数据共享平台的绩效评估	212
7.1.1	评估指标的构建	213
7.1.2	评估方法与评估步骤	216
7.1.3	科学数据共享平台绩效评估的结果	219
7.2	数据保存与共享经济效益识别与评估	231
7.2.1	数据保存与共享成本评估模型	234
7.2.2	数据保存与共享效益识别模型	237
7.2.3	数据保存与共享效益评估模型构建	240
7.2.4	数据保存与共享效益评估的案例	241
第8章	科学数据管理的人才培育	244
8.1	相关文献综述	244
8.1.1	关于对科学数据管理岗位需求的研究	244
8.1.2	关于科学数据管理人才培育的研究	245
8.2	国外图书馆科学数据管理类招聘岗位的调查分析	247
8.2.1	国外图书馆科学数据管理类招聘岗位的调查	247
8.2.2	结论与思考	253
8.3	iSchool 院校科学数据课程及培训计划的调查 与分析	255
8.3.1	iSchool 院校科学数据课程设置的调查与分析	255
8.3.2	教育项目与培训	261
8.3.3	课程内容与岗位职责及能力需求的匹配分析	265
8.4	结论与课程发展的思考	265

8.4.1 培养目标的确立·····	265
8.4.2 培养形式的选择·····	266
附录 科学数据管理与共享政策需求调查问卷·····	268
参考文献·····	274

第 1 章 科学数据管理与共享： 国际组织动向

1.1 国际组织关于科学数据的实践、会议与政策

1.1.1 国际组织关于科学数据共享的行动与项目

(1) 国际科技数据委员会

国际科技数据委员会 (Committee on Data for Science and Technology, CODATA) 是国际科学理事会 (International Council for Science, ICSU) 于 1966 年建立的一个跨学科学术委员会, 致力于提高科技数据的质量、可靠性、管理和可获取性, 推动和鼓励对科学技术有重要价值的可靠的数值数据进行编辑、评价、传播。其目标是: ①提高数据 (尤其是发展中国家) 的质量与可获取性; ②推动数据搜集、组织和使用者之间的国际合作; ③增进国际社会对数据共享重要性的认识; ④考虑数据存取和知识产权问题。CODATA 积极参与国家科学理事会 (ICSU) 开展的一系列重大国际合作项目, 如为全球综合地球观测系统 (Global Earth Observation System of Systems, GEOSS) 和极地信息共享计划 (Polar Information Commons Initiative) 制定广受认可的“数据共享准则”。CODATA 成立一系列任务组开展相关研究与实践活动, 这些任务组由 CODATA 大会评估和决定, 目前正在进行的任务组有 11 个, 大部分是与科研数据有关的项目, 如“数据

风险”(Data at Risk)、“地球和空间科学数据互操作”(Earth and Space Science Data Interoperability)、“数据引用标准与规范”(Data Citation Standards and Practices)、“全球科学信息共享空间”(Global Information Commons for Science)、“关联开放数据推动全球灾害风险研究”(Linked Open Data for Global Disaster Risk Research)、“发展中国家科技数据的保存和获取”(Preservation of and Access to Scientific and Technical Data in/for/with Developing Countries)等。

此外,CODATA 出版了一系列图书和专著,定期发布生物、环境、地质、材料、热力等各学科的公报,还发行了免费的同行评审电子期刊《CODATA 数据科学期刊》(CODATA *Data Science Journal*)^①等。

(2) 联合国教科文组织

联合国教科文组织(United Nations Educational, Scientific, and Cultural Organization, UNESCO)于2007年启动了“促进发展中国家科学数据共享与应用全球联盟”(Global Alliance for Enhancing Access to and Application of Scientific Data in Developing Countries, e-SDDC)计划。该项目是经联合国全球信息通信技术与发展联盟(Global Alliance for ICT and Development, UN GAID)批准,由中国科学院领衔,国际科学技术数据委员会与发展中国家科学数据保护与共享工作组、巴西国家环境参考信息中心、南非国家基金会、美国哥伦比亚大学国际地球科学信息网络中心、中国科学技术协会、中国互联网协会等组织共同参与,目的是在UN GAID的框架下,在全球范围内建立一个以应用为导向,学术界、政界、企业界相关单位通过双赢机制联合起来的合作平台,促进发展中国家科研数据的广泛共享与应用,发挥科研数据在促进社会发展应用中的增值作用^②。该项计划由5项具体实施行动组成,分别是:组织智囊团网络、建设科研数据中心网、技术转让和技术援助网、在线培训网和应用示范网。目

^① CODATA Data Science Journal [EB/OL]. [2015-02-01]. <https://www.jstage.jst.go.jp/browse/dsj>.

^② 科学网. “促进发展中国家科学数据共享与应用全球联盟”计划启动 [EB/OL]. [2015-02-10]. <http://news.sciencenet.cn/htmlnews/2007519132078179754.html>.



前,项目先后成功召开了巴西圣保罗会议、南非普里托利亚会议和美国华盛顿会议,相关活动正按计划进行。

1.1.2 国际组织有关科学数据共享的会议

(1) 国际科技数据委员会

国际科技数据委员会(CODATA)自1968年起每两年举办一次国际学术会议,邀请世界各领域的科学家共同探讨科学数据领域面临的问题和挑战,并对任务组的工作进行评审,决定是否继续进行。其中,第20届CODATA国际学术会议于2006年10月在北京召开,会议的主题是“信息社会的科学数据与知识”,对数字鸿沟、数据归档和存储、数据共享政策与法规、数据可视化与多媒体技术、数据挖掘与知识发现、国际极地年(IPY)以及科学数据在教育、经济、环境、社会和科学研究中的应用等问题进行了广泛探讨^①,会议成立了“数据政策咨询委员会”(Data Policy Advisory Committee),负责追踪国际数据政策议题并为CODATA提供建议。为了缩小发达国家和发展中国家的数字鸿沟,本次CODATA会议启动了“科学数据跨越数字鸿沟”(Scientific Data across the Digital Divide Program)项目,目的是改善发展中国家收集、获取、管理和利用科学数据能力,促进可持续发展。最近一届(第24届)CODATA国际学术会议(SciDataCon 2014)于2014年11月在新德里举行,会议以“面向全球可持续发展的数据共享与整合”(Data Sharing and Integration for Global Sustainability)为主题,围绕政府数据开放与智慧城市、数据服务的可持续性,数据共享与引用,数据基础设施与数据挖掘,开放科学的实施策略,跨学科间的数据互操作、风险数据及其恢复技术等问题进行了研讨^②。

(2) 国际科学技术信息委员会

国际科学技术信息委员会(International Council for Scientific and

^① 第二十八届国际科技数据委员会学术会议综述[J]. 中国科学院院刊, 2006(6):510-512.

^② CODATA. SciDataCon2014[EB/OL].[2015-02-06]. <http://www.scidata-con2014.org/conference-programme>.

Technical Information, ICSTI)也是由国际科学理事会(ICSU)建立的,是一家涉及科技信息的公益性非政府国际科学协会组织,致力于推动科学交流合作、提高科研工作成效,为世界科技信息领域内的交流提供平台和纽带。目前该组织在全球拥有包括各国科技信息机构在内的40多个会员单位。

ICSTI每年定期召开夏季大会,来自各国科技信息机构的专家学者围绕会议主题进行学术交流讨论。其中,2011年的年度会议由中国科学技术信息研究所承办,会议主题为“迈向知识服务”,对信息资源和服务的开发、知识组织和发现、网络环境下的数字公平和社会发展、数字环境下的信息服务整合、数字环境下的知识产权等问题进行了深入讨论。

(3) 国际图联

随着国际社会对科学数据越来越广泛的关注,国际图联(International Federation of Library Associations and Institutions, IFLA)也开始重视这个问题。2012年8月在芬兰首都赫尔辛基举行的第78届IFLA世界图书馆与信息大会设置了一个专门分会探讨与科研数据有关的问题。该分会以“科学与技术图书馆在数据监管、存取与保存中的角色:一个国际视角”(The Role of Libraries in Data Curation, Access and Preservation: An International Perspective—Science and Technology Libraries)^①为主题,共收到来自欧洲和北美地区的6篇论文,对图书馆在科研数据监管与保存中的角色、图书馆员对开展数据服务的态度与能力胜任、加拿大国家科研数据计划、图书馆在数据交换中的角色等问题进行探讨。

(4) 欧洲科学基金会

欧洲科学基金会(European Science Foundation, ESF)成立于1974年,是由隶属于30多个国家的77个科学研究委员会、科学院、研究院和其他科研资助机构组成的独立性非政府组织,旨在鼓励、支

^① IFLA. IFLA World Library and Information Congress—78th IFLA General Conference and Assembly [EB/OL]. [2015-02-06]. <http://conference.ifla.org/ifla78/programme-and-proceedings-full-printable>.

持和帮助成员在科学研究方面的交流与合作,并就欧洲科学政策的重大战略性课题组织成员开展共同研究。欧洲科学基金会重视数据对科学研究的重要意义,于2007年和德国科学基金会共同举办了“研究数据共享中的共同责任”(Shared Responsibilities in Sharing Research Data)研讨会,指出研究数据共享对提高科研效率和确保研究完整性有着重要的意义,科研资助机构在促进研究数据共享中应扮演更加关键的角色,通过鼓励研究数据的开放共享、优先资助那些制定良好数据共享计划的项目、制定有效的数据共享政策等措施来促进研究数据的共享^①。

1.1.3 国际组织有关科学数据共享的政策与声明

(1) 联合国教科文组织

UNESCO于2004年发布了《开发和推广政府公共领域信息政策指南》(*Policy Guidelines for the Development and Promotion of Governmental Public Domain Information*)^②(以下简称《UNESCO指南》)。《UNESCO指南》支持研究数据和其他公共信息的开放传播,指出:公共资助的科研数据和公共领域未加保护的事实信息是基础研究的重要支柱,通过公共部门资助的科学技术信息的开放和高效获取,遵守国家安全控制和他人来自保密义务的权利、知识产权和隐私保护等以推进卓越研究和公共研究的有效利用与发展基金。

《UNESCO指南》将在国家与国际层面的决策和政策制定者中得到广泛应用。然而,《UNESCO指南》仅仅是一个报告,而不是一个规定性或规范性的工具。此外,UNESCO于2007年发布的《开发与促进开放获取的政策指南》(*Policy Guidelines for the Development*

^① ESF. Shared Responsibilities in Sharing Research Data; Policies and Partnerships [EB/OL]. [2015-02-06]. http://www.esf.org/fileadmin/Public_documents/Publications/SharingData_01.pdf.

^② UNESCO. Policy Guidelines for the Development and Promotion of Governmental Public Domain Information [EB/OL]. [2015-02-06]. <http://unesdoc.unesco.org/images/0013/001373/137363eo.pdf>.

and Promotion of Open Access^①)将研究数据纳入开放获取的范围,强调科学数据同其他类型资料一样,都应免费、开放获取。

(2) 世界经济合作与发展组织

世界经济合作与发展组织(Organization for Economic Cooperation and Development, OECD)于2004年1月30日通过了由包括中国在内的30多个OECD成员国共同签署的《公共资助的研究数据开放存取宣言》(*Declaration on Access to Research Data from Public Funding*) (以下简称《宣言》)^②。《宣言》强调国际间的数据、信息和知识交流,将极大地推动科研发展与创新,数据的开放获取和无限制利用将推动科学进展和促进研究者的培训,开放获取有利于公共投资收集的数据价值最大化等理念。《宣言》成员国承诺在推动公共资助的研究数据获取过程中将参照“开放”“透明”“法律一致性”“保护知识产权”等10条原则。

为了进一步明确和强化《宣言》中原则的指导作用,OECD于2007年4月发布了《OECD关于公共资助的研究数据获取的原则与指南》(*OECD Principles and Guidelines for Access to Research Data from Public Funding*) (以下简称《OECD原则与指南》)^③。《OECD原则与指南》指出:“成员国应考虑并应用本文件中推荐的公共资助的研究数据获取的原则与指南。每个成员国都应为研究数据的访问、利用和管理制定政策和良好的措施。”其目标主要是:推进开放的文化和数据共享,提高对潜在成本和利益的认识;强调在科技政策中考虑数据访问与共享法规和措施的必要性,提供一个可操作的原则和优秀实践的框架,提供宽广政策推荐以改进国际数据获取与共享。

① UNESCO. Policy Guidelines for the Development and Promotion of Open Access [EB/OL]. [2015-02-06]. <http://unesdoc.unesco.org/images/0021/002158/215863e.pdf>.

② OECD. Declaration on Access to Research Data from Public Funding [EB/OL]. [2015-02-06]. <http://archiv.twoday.net/stories/133859/>.

③ OECD. OECD Principles and Guidelines for Access to Research Data from Public Funding [EB/OL]. [2015-02-06]. <http://www.oecd.org/dataoecd/9/61/38500813.pdf>.