

近十五年来科学数据管理领域国际组织实践研究^{*}

黄如花 周志峰

摘要 本文介绍了近十五年来联合国、欧盟、经济合作与发展组织、国际科学技术数据委员会、研究数据联盟、国际图联在科学数据管理领域的主要相关活动,并提出应从宏观、中观和微观三个层面,借鉴国际组织关于科学数据管理的实践成果与经验,以期推动我国相关实践活动的开展。参考文献 51。

关键词 科学数据 科研数据 科学数据管理 国际组织

Study on Practices of International Organizations in the Field of Scientific Data Management during the Last 15 Years

Huang Ruhua Zhou Zhifeng

Abstract: This paper introduces the major related events of international organizations in the field of scientific data management during the last 15 years including UN, EU, OECD, CODATA, RDA and IFLA. It also proposes that we should draw on the practical results and experiences of scientific data management in international organizations on micro-level, meso-level and macro-level in order to promote the practices of scientific data management in China. 51 refs.

Keywords: Scientific Data; Research Data; Scientific Data Management; International Organization

科学数据是重要的公共信息资源,也是宝贵的科研资产。2013 年 1 月,国务院发布《“十二五”国家自主创新能力建设规划》,提出“构建科技资源从数据获取、存储、处理、挖掘到开放共享的完整信息服务链,建设集中与分散相结合的国家科学数据中心群,形成国家科学数据分级分类共享服务体系”^[1];2015 年 8 月,国务院出台《促进大数据发展行动纲要》,进一步提出“发展科学大数据,积极推动由国家公共财政支持的公益性科研活动获取和产生的科学数据逐步开放共享,构建科学大数据国家重大基础设施,实现对国家重要科技数据的权威汇集、长期保存、集成管理和全面共享。面向经济社会发展需求,发展科学大数据应用服务中心,支持解决经济社会发展和国家安全重大问题”^[2]。本文拟汇总近十五年来

国际组织在科学数据管理领域的实践活动,希望为我国“科学数据共享工程”和“中科院数据应用环境建设与服务”等国家级科学数据管理项目效率的提高与优化以及科研发展的数据驱动创新提供借鉴。

1 联合国:影响最广的科学数据管理活动的推动者

作为世界上最重要的国际组织,联合国关于科学数据管理的态度和行动已在全球范围内产生重大影响。

1.1 发布宣言促进和规范科学数据的共享与利用

2003 年 10 月,联合国教科文组织 (United Nations Educational, Scientific and Cultural Organi-

^{*} 本文系武汉大学人文社会科学自主科研项目“促进我国数据开放共享的对策研究”研究成果之一,并得到“中央高校基本科研业务费专项资金”资助。

ization, 简称 UNESCO) 通过《国际人类基因数据宣言》(*International Declaration on Human Genetic Data*), 倡议以伦理道德准则规范人类基因科学数据的采集、处理、储存及使用^[3]。

2012 年 9 月, UNESCO 在“数字化时代的世界记忆: 数字化和保存”国际会议上通过《温哥华宣言》(*UNESCO/UBC Vancouver Declaration*), 指出“为了确保能够在未来获取今天创造的数据和产生的记录、避免数字失忆, 人类面临着巨大的挑战”; 强调《世界人权宣言》(*The Universal Declaration of Human Rights*) 应保证每个人获取信息的权利, 包括数字格式的信息; 应该通过建立相应的国家政策来保证信息权利, 促进数据的开放^[4]。

1.2 召开重要会议讨论科学数据管理议题

2005 年, UNESCO 与国际图联 (*International Federation of Library Associations and Institutions*, 简称 IFLA) 在合作主办的信息社会世界峰会 (*The World Summit on the Information Society*, 简称 WSIS) 突尼斯阶段会议中, 针对科学数据共享问题进行了专门的讨论^[5]。

2009 年, UNESCO、国际电信联盟 (*International Telecommunications Union*, 简称 ITU)、联合国贸易及发展大会 (*United Nations Conference on Trade and Development*, 简称 UNCTAD) 和联合国开发计划署 (*United Nations Development Programme*, 简称 UNDP) 等共同组织 WSIS 论坛, 召开“WSIS 存取行动线与数字科研行动线”会议, 重点讨论了促进科学数据开放存取和知识可持续发展的行动方针^[6]。2011 年 5 月, WSIS 论坛“Action Lines C3 & C7”会议召开, 议题包括开放存取出版、研究出版物与研究数据的关联^[7]。

在世界首脑级峰会上讨论科学数据管理问题, 表明国际社会已经认识到科学数据充分、有序共享是信息社会发展的必然和需要, 但实现这种共识需要科学数据的各利益相关者一起以协商和互惠为总原则, 共同构架起加速 21 世纪科技

发展的数据平台。

1.3 促成科学数据管理领域的国际合作

2005 年, UNESCO 牵头发布了《全球科学信息共有倡议》(*The Global Information Commons for Science Initiative*), 激励人们通过创新模式促进科学数据的传播和合作利用, 为共享经验、建立合作、制定开发原则与指南以及最佳实践推广提供一个全球化平台^[8]。

2007 年, “促进发展中国家科学数据共享与应用全球联盟 (UN e-SDDC)” 计划获得联合国全球信息通讯技术与发展联盟 (*United Nations Global Alliance for Information and Communication Technologies and Development*, 简称 UN GAID) 批准。该计划由中国科学院牵头, 国际科技数据委员会 (*The Committee on Data for Science and Technology*, 简称 CODATA)、美国哥伦比亚大学国际地球科学信息网络中心、南非国家基金、巴西国家环境参考信息中心等参与其中, 其目标是在科技界尽量缩小发展中国家与发达国家之间的数字鸿沟, 提出一个全球性的促进发展中国家科学数据共享与应用的解决方案^[9]。

1.4 建立开放数据门户

联合国经济与社会事务署 (*The United Nations Department of Economic and Social Affairs*, 简称 UNDESA) 统计司推出的面向全球用户的数据门户——UNDATA, 截止到 2015 年 12 月 29 日, 已收集了 34 个数据库、6000 多万条统计数据记录, 涉及能源、环境、农业、人口、贸易、工业、教育、旅游、犯罪等方面, 便于用户进行一站式检索与下载联合国系统的各种统计数据。而且, UNDATA 为用户提供直接从各国相关数据库中进行搜索的服务, 数据源的权威性能够得到很好的保证^[10]。据该网站的使用情况统计, 自 2008 年 4 月到 2014 年 12 月, 共有来自世界 243 个国家或地区的 1870 万次访问, 因不断提升用户服务质量, UNDATA 网站获得 2010 年“联合国 21 世纪奖 (UN21 Award)”^[11]。

1.5 发布科学数据管理的标准和工具

联合国粮农组织 (Food and Agricultural Organization of the United Nations, 简称 FAO) 负责管理农业信息管理标准 (Agriculture Information Management Standards, 简称 AIMS) 的门户网站, 致力于传播农业信息管理的标准和良好规范^[12]。AIMS 为农业研究与发展信息共享体系 (Coherence in Information for Agricultural Research for Development, 简称 CIARD) 开展国际合作计划提供支持, 促进了信息和服务的共享。在 AIMS 的框架下, FAO 现已开发了 AGROVOC、AgroTagger 等词表, 农业元数据要素集 (The Agricultural Metadata Element Set)、有意义的书目元数据 (Meaningful Bibliographic Metadata) 等元数据标准, 以及 AgriDrupal、农业海洋数字式空间 (AgriOceanDSpace) 和农业元数据生成器 (AgriMetaMaker) 等工具, 并将 AGROVOC 等词表发布成为关联数据 (Linked data)。这些词表、标准、工具为农业信息, 尤其是农业科学数据的发布、互操作和再次利用提供了便利。

1.6 推动科学数据的共享和应用

2015 年 10 月, 在《粮食和农业植物遗传资源国际条约》(The International Treaty on Plant Genetic Resources for Food and Agriculture, 简称 ITPGREA) 框架下, FAO 宣布国际水稻研究所 (International Rice Research Institute, 简称 IRRI) 提供的超过 3000 个不同品种的水稻的基因组序列科学数据将在全球范围内共享, 以促进相关物种研究, 确保全球未来粮食安全^[13]。在联合国全球地理信息管理专家委员会 (United Nations Global Geospatial Information Management, 简称 UN-GGIM) 的推动下, 2015 年 2 月, 第 69 届联合国大会第 80 次会议通过了“促进可持续发展的全球大地测量参考框架”, 敦促共享地理空间信息与数据, 以造福人类。地理空间相关科学数据能够支持可持续发展政策的制定、气候变化的监测和自然灾害的管理, 也广泛应用于交通、农业和建筑等各行业^[14]。

2 欧盟: 构建欧洲科学数据管理的行动框架和发展愿景

欧盟 (European Union, 简称 EU) 关于科学数据管理的动向对欧洲国家产生了直接影响, 也在国际范围内起到了示范和引导作用。由于欧盟在一定程度上具有准政府的职能, 因此它制定的关于科学数据管理的法律和政策等都具有直接的执行效力。

2.1 逐步完善科学数据共享相关法律法规

在科学数据共享的立法和政策制定上, 欧盟确定了指导思想和最终目标。2002 年, 在布加勒斯特泛欧大会上, 欧盟发布《布加勒斯特宣言——迈向信息社会: 原则、战略和优先行动》, 提出了公共部门持有信息的公益性共享原则和指导思想。该宣言中有关公共数据共享的内容, 为欧盟制定科学数据共享法律规则提供了基本原则和思想基础^[15]。

在与科学数据共享相关的法律法规方面, 欧盟除了制定具有指导意义的《欧盟条约》和各国的信息公开法之外, 还涉及了科学数据共享保障体系相关的众多领域, 如基础设施、技术平台、数据保护、网络和信息安全、知识产权、财政援助等。

2.2 积极开展科学数据管理相关研究

2003 年 12 月, 信息基础设施咨询工作组 (e-Infrastructure Reflection Group, 简称 e-IRG) 成立。e-IRG 作为 EU 的高层次战略研究组织, 旨在有效地建设欧洲科研基础设施与开展科学数据管理, 其任务是在政策、咨询和监督层面上为 e-Infrastructure 的战略提出建议。e-IRG 成员们认识到数据管理在未来研究中的重要性, 专门成立了数据管理任务组 (e-IRG Data Management Task Force)。2009 年 12 月, e-IRG 发布《数据管理报告》, 对数据管理计划、元数据及其质量、数据管理的互操作等进行了分析^[16]。2012 年 1 月, e-IRG 发布研究报告, 为实现数字议程目标所需开展的行动提供指南, 其中数据管理的互操作问题是优先关注的领域之一; 强调为确保科学数据的

跨学科利用,应支持鼓励相应项目的开展;并特别提出以下建议:鼓励开发非特定学科的框架和信息架构,以实现数据的互操作交换;支持科研团体,确定他们对语义互操作性的需求和拟开展的活动;在分析现有接口和协议标准利用现状及局限性的基础上,将实现互操作的接口和协议标准化作为优先领域^[17]。

2.3 构想科学数据信息化的发展愿景

2.3.1 2030 年的发展愿景

2008 年,EU 就已提出科学数据长期保存计划。2010 年 10 月,EU 数字议程发布的报告《弄潮:欧洲如何在科学数据潮流中成为赢家》勾画出了 2030 年科学数据信息化基础设施的发展愿景——支持对数据的无缝访问、利用和再利用,描述了与科学数据访问、管理、保存相关的挑战,制定了实现此愿景所需要的战略和 6 项行动建议^[18]。

2.3.2 GRDI2020 愿景与建议

GRDI2020 项目由 EU“第七框架计划”(FP7)资助。2011 年 1 月,GRDI2020 发布报告《全球科学数据基础设施:重大数据挑战》,阐述了构建全球科学数据基础设施所面临的主要挑战和亟需解决的问题,进而为实现全球科学数据基础设施的愿景提出了相关建议:支持数据密集型、多学科与跨学科的研究;开发新型数据工具;支持构建开放链接的数据空间;开发和实施与数据、元数据等相关的正式模型和查询语言;支持科学数据与文献之间的互操作等^[19]。

2.3.3 2020 年行动框架展望

2010 年,EU 启动了“欧洲 2020 战略”,作为落实该战略的“第七框架计划”已于 2013 年底结束,2014 年又启动“地平线 2020——研究与创新框架计划”(Horizon 2020 The EU Framework Programme for Research and Innovation)^[20]。在数据驱动科学和开放获取政策与倡议的影响下,EU 发布了公众咨询报告《研究数据电子基础设施:展望 2020 行动框架》(Research Data e-Infrastructures: Framework for Action in Horizon 2020)^[21],提出通过

相关利益方(如研究机构、研究人员、图书馆、数据中心、出版商等)的积极参与,开发具有开放性、互操作性的研究数据电子基础设施。该行动框架认识到不断变革的研究本质,以及使整个科研生态系统中的相关方都参与到开发与整合研究数据电子基础设施过程中的重要性。为支持“开放式数字驱动的科学”,该行动框架列举了 7 个行动领域:科学研究共同体支持数据服务,开放存取的基础设施,研究数据的存储、管理和保存,研究数据发现与溯源,面向全球化的研究数据电子基础设施,电子基础设施的身份验证和授权,与研究数据相关的技能和新职业等。

2013 年 3 月,欧洲研究图书馆协会(Ligue des Bibliothèques Européennes de Recherche,简称 LIBER)、欧洲研究开放获取基础设施项目(Open Access Infrastructure for Research in Europe,简称 OpenAIRE)和开放获取知识库联盟(Confederation of Open Access Repositories,简称 COAR)联合响应该行动框架,在赞同其举措的同时,也提出了相关建议^[22],如:为支持研究数据的发现、导航与利用,构建一个能使所有利益相关者参与、普遍适用于所有行动领域的实施方案;关注研究机构与研究活动的长尾效应;研究机构与数据中心之间应建立清晰的工作流程;解决人文学科研究数据和电子技术应用问题等。

2015 年 10 月,EU 发布《科学出版物与研究数据开放存取指南(第二版)》,计划实施开放科学数据试点(Open Research Data Pilot)^[23]工作。

3 经济合作与发展组织:关注科学数据开放共享与数据驱动创新

经济合作与发展组织(Organization for Economic Cooperation and Development,简称 OECD)旨在共同应对全球化带来的经济、社会和政府治理等方面的挑战。在应对科学数据管理领域的挑战时,OECD 也表明了自己的看法和应对态度。

3.1 推进公共领域科学数据的共享与获取

十余年来,OECD 持续关注公共领域的科学

数据共享与获取问题。OECD 所发布的宣言对科学数据的共享与获取做出了原则性的阐述,虽然没有产生直接的法定约束力,但在国际范围内特别是其成员国内,对科学数据管理相关政策和法规的制定产生了深远影响。

2004 年 1 月,OECD 在巴黎发表《公共基金资助研究数据存取宣言》(*Declaration on Access to Research Data from Public Funding*),包括中国在内的 30 个国家承诺公共基金资助的研究项目的数据存取应向以下目标努力,即开放、透明、法律环境、正式责任、专业、保护智慧财产、互操作性、品质与安全、效率、问责等^[24]。

2006 年 12 月,为对上述宣言提供具体行为准则,OECD 颁布的《公共资金资助的研究数据获取原则与指南》(*OECD Principles and Guidelines for Access to Research Data from Public Funding*)要求其成员国在制定并完善与科学数据共享相关的法律和政策时,将这些原则应用其中,并增加或修订了弹性、长期性、品质、安全等原则^[25]。

2008 年 6 月,“网络经济的未来”部长级会议在韩国首尔举行,OECD 成员国签署《网络经济的未来:首尔宣言》(*The Seoul Declaration for the Future of the Internet Economy*),宣言提出要通过政策制定来维护科研与创新的开放环境,使公共部门包括科学数据在内的信息和内容能以更加广泛的形式获取”^[26]。

3.2 强调科学研究领域的数据驱动创新 (Data-Driven Innovation)

OECD 在《科学技术与工业展望 2012》报告中提到,科学的发展越来越依赖于数据驱动,但科学数据的开放却受到行政、法律等的束缚^[27]。2013 年,OECD 在《探索数据驱动创新作为一种新的增长源泉》报告中指出,数据已成为打造显著竞争优势、驱动创新和可持续发展的核心要素^[28]。

2015 年 10 月,OECD 发布报告《使开放科学成为现实》^[29],认为网络与在线平台为科学研究中大型数据集的组织与发布提供了新的机遇,使

得科学发展越来越多地被数据所驱动;但在各国的开放科学政策方面,激励措施往往是对开放获取出版成本的资助,而很少涉及对科研人员在开放数据行动方面的补偿机制。报告还指出:与科学文献的开放获取相比,目前各国关于开放数据的政策不够完善;需要更好的激励机制促进科研人员之间的数据共享;科学界需进一步提升与数据相关的技能;国际和国家层面都需要更加清晰的关于文献共享和数据重用的立法。2015 年 11 月,OECD 在《大数据促进创新以振兴经济和提高人类福祉》的报告中提出^[30]:推动数据驱动的科学研究,建立跨国家和自由流动的数据组织,利用大数据来提高效率,促进知识共享;同时,数据共享应制定相关的法律保护条例,如专利保护、知识产权等,使知识信息共享有法可依。

4 国际科学技术数据委员会:提供技术指导

国际科学技术数据委员会(CODATA)作为国际科学理事会(International Council for Science Unions, 简称 ICSU)下属的跨学科国际组织,是促进科学数据发展的专门机构,关注科学技术各个领域的实验测量、观察和数据计算,尤其关注不同学科所共有的数据管理问题及数据在其产生学科领域之外的应用。CODATA 的主要作用是促进和培育科学数据管理方面的广泛合作与交流,并提供技术指导。

4.1 组织针对特定数据问题的任务组与工作组

2014 年,经 CODATA 大会批准和更新,目前共设置了针对特定数据问题的 2 个委员会(Committees)、10 个任务组(Task Groups)、5 个工作组(Working Groups),同时还参与了地球观测组织(Group on Earth Observations, 简称 GEO)、研究数据联盟(Research Data Alliance, 简称 RDA)、世界数据系统(World Data System, 简称 WDS)设立的 3 个工作组^[31]。

4.1.1 全面推动数据科学的发展

数据科学已经逐步发展成为一门具有明确

研究对象、独特研究方法、广泛研究领域和应用领域的交叉学科。早在 1969 年, CODATA 就成立了“基本物理常数”任务组(Fundamental Physical Constants), 测定基本物理常数; 2010 年, 成立了“数据引用标准与实践”任务组(Data Citation Standards and Practices)和“数据风险管理”任务组(Data at Risk), 关注科学数据的保存和引用标准问题。

CODATA 任务组不断完善数据科学基础理论, 特别是在科学数据基本概念标准、科学数据引文标准和海量数据风险管理标准方面, 其成果为科学数据相关国际标准的建立奠定了基础。

4.1.2 促进全球科学数据的共用与共享

CODATA 倡导科学数据是人类共同财产的理念, 并发起了科学数据共享全球战略和政策研究, 组织全球科学数据共享管理的国际联合行动计划。CODATA 长期发展战略将全球科学数据管理作为其重要任务之一, 涉及领域包括人体健康、材料科学、地球科学、社会科学以及综合性科学等。

(1) 2004 年, CODATA 成立“人体测量数据和工程学”任务组(Anthropometric Data and Engineering), 促进国际范围内人体测量数据的协作收集、共享利用。

(2) 2008 年, CODATA 设立“地球与空间数据互操作”任务组(Earth and Space Science Data Interoperability), 促进地球科学数据和元数据交换以及全球整合。

(3) 2010 年, CODATA 设立“全球道路数据开发”任务组(Global Roads Data Development), 加强国际科学数据的使用, 通过开放访问向国际社会提供更好的道路数据, 以造福社会。

(4) 2012 年, CODATA 成立“推进微生物学信息学”任务组(Advancing Informatics for Microbiology, 简称 TG-AIM), 促进和维持微生物研究数据的共用与共享, 特别是数据访问和使用的互操作性。

(5) 2012 年, CODATA 设立“全球灾害风险研究关联开放数据”任务组(Linked Open Data for

Global Disaster Risk Research, 简称 LODGD), 促进全球灾害风险数据的互联共享。

(6) 2014 年, CODATA 成立“互操作数据出版物”任务组(Interoperable Data Publications), 促进科学数据的互操作和数据发布。

4.1.3 寻求与其他相关组织的深入合作

CODATA 设立工作组的目的旨在探索解决战略发展过程中一些短期问题的具体方案。CODATA 与 RDA、GEO、WDS 等国际上与数据相关的组织有广泛合作, 其中设立工作组是重要的形式。联合工作组主要有两种:

一是 CODATA 处于主导地位, 如与 RDA 设立的 3 个工作组, 即研究数据法律互操作兴趣组(CODATA-RDA Interest Group on Legal Interoperability of Research Data)、研究数据科学暑期学校工作组(CODATA-RDA Working Group on Research Data Science Summer Schools)、国际材料资源注册工作组(CODATA-RDA Working Group on International Materials Resource Registry), 与新材料与标准凡尔赛公约组织(Versailles Project on Advanced Materials and Standards, 简称 VAMAS)设立纳米材料描述联合工作组(CODATA/VAMAS Joint Working Group on the Description of Nanomaterials)。

二是 CODATA 参与到其他组织的工作组中, 如 GEO 的数据共享工作组(GEO Data Sharing Working Group)和数据管理原则工作组(GEO Data Management Principles Task Force)、RDA-WDS 的数据中心成本回收兴趣组(RDA-WDS Interest Group on Cost Recovery for Data Centres)。

4.2 提供国际科学数据共享的研究与实践指导

4.2.1 对国际科学数据管理进行长期规划

进入 21 世纪以来, 随着 E-science 和数据驱动型科学研究的发展, CODATA 开始注重对国际科学数据管理进行战略性的规划。2005 年初, 为了规划数据管理的发展前景, CODATA 成立战略规划委员会, 并开始起草战略计划。2008 年 10

月,在乌克兰基辅召开的 CODATA 第 26 届大会上,CODATA 正式批准自成立以来的第一个战略规划《战略规划(2006—2012)》^[32];2012 年 10 月,CODATA 大会通过了“2013—2014 年科学议程”^[33]。

2013 年 3 月,CODATA 发布《战略规划(2013—2018)》最终版^[34]。推出三大计划:

(1)发挥 CODATA 在科学数据政策、数据科学的体制框架以及数据管理方面的领导作用,开展广泛的合作,确定与科学数据未来发展最相关的关键政策和体制框架,与科学界重要的政策目标和战略相结合,并加以实现;

(2)明确 CODATA 在数据科学与数据管理中可以发挥其实质性影响力的关键前沿领域以及重要的跨学科应用领域;

(3)帮助 ICSU 积极主动地解决科学数据管理问题,包括基础设施、政策、流程和资源等。

《战略规划(2013—2018)》还提出了 CODATA 的愿景和发展方向,包括扩大 CODATA 的作用、厘清 CODATA 与 WDS 之间的关系、加强 CODATA 会员的建设、结合战略目标拓展 CODATA 的服务范围、加强 CODATA 的财务规划、实现可持续发展等。

4.2.2 组织学术交流活动

(1)定期召开 CODATA 大会

CODATA 大会和全会每两年举行一次,是 CODATA 的主要活动之一,CODATA 大会被誉为“科技数据领域的联合国大会”,是全球科技数据领域交流合作的重要平台^[35]。

2012 年 10 月 28 日至 11 月 2 日,第 23 届 CODATA 大会暨第 28 届 CODATA 全会在台北市召开,会议以“开放数据与信息:变化中的地球(Open Data and Information for a Changing Planet)”为主题。会议围绕国家和地区的数据政策、数据保藏与管理、数据共享与引用、开放数据环境的概念与应用、数据密集型科学问题、跨学科间的数据应用、灾害数据的处理技术与基础设施建设等问题进行了研讨,在 CODATA《战略规划(2013—2018)》中,对原有和新申请任务组进行

了讨论,最后批准设立 12 个任务组。

2014 年 11 月 2—5 日,CODATA 与世界数据系统(ICSU-World Data System)、印度科学院共同组织了国际科学数据会议(SciDataCon),会议是 CODATA 大会的升级版,主题涉及发掘数据服务于全球可持续发展、全球数据挑战及开放数据、数据共享整合与集成对科学、社会、经济和政策的影响等。

2016 年 9 月 11—17 日,SciDataCon 2016 将在美国科罗拉多州的丹佛召开,主题为“推进科学研究领域的数据前沿”(Advancing the Frontiers of Data in Research),围绕数据革命和数据驱动的科学研

(2)开设学术研讨会、培训班

每年 CODATA 的任务组、工作组和成员都会组织学术研讨会,也举办围绕科学数据主题的国际研讨班,特别是为发展中国家培训科学数据管理与服务的专门人才^[36]。

2014 年 6 月,CODATA 与中科院联合举办“发展中国家科学大数据国际培训班”,为发展中国家的科研与数据工作者提供科学数据管理、共享与服务、大数据处理等方面的知识与技能;同期,“大数据与国际科学计划研讨会”召开,审议关于大数据与科学发现的声明,提出大数据促进科学发展的建议与行动纲领。2014 年 8 月,CODATA 的“发展中国家科技数据保藏与共享”任务组在肯尼亚首都内罗毕举办开放数据研讨班;同时,召开了主题为“科学开放数据与发展中国家的可持续发展”的学术研讨会。

2015 年 3 月,CODATA 与印度统计研究所合作,在班加罗尔举行关于大数据的研讨班,培养初、中级专业的大数据管理和开发专业技术人才。2015 年 9 月,CODATA 的成员印尼科学研究院在雅加达举行大数据管理与分析的研讨班。

(3)筹备国际数据周活动

2016 年 9 月 11—17 日,CODATA 将举行国际数据周(International Data Week)活动^[37],主题是“从大数据到开放数据:动员数据革命”(From Big Data to

Open Data: Mobilizing the Data Revolution), 活动主要包括第八次研究数据联盟 (Research Data Alliance, 简称 RDA) 全体会议、SciDataCon 2016 会议以及国际数据论坛三个部分。国际数据周活动将汇集世界范围内的数据科学家、研究人员、行业领袖、企业家、政策制定者和数据管理员等, 一起探讨如何更好地通过数据驱动的研究和创新来提高社会的知识和利益。

4.2.3 出版《数据科学杂志》等刊物

为了给科学数据管理的研究者与从业者提供信息交流的平台, 推动相关的理论研究和实践探索, 出版与科学数据相关的研究报告、会议录等各类出版物, CODATA 不断丰富自身职能, 主办《数据科学杂志》(CODATA Data Science Journal)^[38]。《数据科学杂志》的创立标志着数据科学作为一门独立的学科正式诞生^[39]。

4.3 促进科学数据的国际共享

4.3.1 协调国际数据项目

2011 年, CODATA 实施“手拉手合作伙伴计划”(Hand in Hand Program), 旨在促进国际项目之间的合作, 更充分地使用已有数据资源, 有效地减少资料收集、数据采集等重复劳动及由此产生的相应费用, 帮助科学家提高科研成果产出, 加速科研成果转化, 消减数字鸿沟。目前, 该计划已经启动与“灾害风险综合研究计划”(Integrated Research on Disaster Risk Program) 等的合作^[40]。

4.3.2 指导发展中国家的实践

2002 年, 为了关注发展中国家科学数据管理的实践活动, CODATA 成立了“发展中国家科技数据保藏与共享”任务组(Preservation of and Access to Scientific and Technical Data in/for/with Developing Countries, 简称 PASTD)^[41]。PASTD 促进了对发展中国家关于科学数据长期保存、归档和永久访问需求的更深入了解, 推动了相关技术标准 and 政策的制定。PASTD 自成立以来, 先后得到 UNESCO、联合国经社事务部信息通讯与发展全球联盟 (UNDESA GAID)、国际科学院组织

(IAP)、全球变化亚太网络组织 (APN), 以及美国、中国、南非、巴西等国家基金会的支持, 在亚洲、非洲、拉丁美洲先后召开过一系列科学研讨会和数据处理与应用技术培训。PASTD 在新阶段的发展计划中, 将发展中国家科学数据共享与应用能力建设作为重点, 这与联合国促进发展中国家科学数据共享与应用全球联盟 (UN GAID e-SDDC) 的计划相互结合、互为补充。

2012 年 7 月, CODATA 举办了“发展中国家科学数据管理和可持续发展培训国际研讨会”, 聚焦于科学数据生命周期管理以及可持续使用^[42]。该会议以中科院在科技数据管理与共享方面的理论和实践经验为基础, 为发展中国家的科技人员与数据工作者提供系统全面的培训。培训内容包括科技数据资源规划、标准规范、管理的技术手段及典型案例等。培训的开展为建立发展中国家科技数据共享网络、促进发展中国家在国际科技合作项目上开展实质性深入合作奠定了基础^[43]。

5 研究数据联盟: 致力于推动全球科学数据的开放与共享

2013 年, 在 EU、澳大利亚、美国等的资助下, 研究数据联盟 (Research Data Alliance, 简称 RDA) 成立, 目标是希望通过基础设施、标准、政策和实践等方面的发展促进研究数据的共享与交流、使用与复用, 推进全球数据驱动的创新与发展^[47]。截止到 2016 年 1 月, RDA 已有超过 3600 个成员, 遍布全球 105 个国家。

5.1 频繁召开会议推进工作开展

RDA 全会是工作会议, 每半年召开一次, 会议汇集数据科学家、图书馆员、计算机科学家和其他领域科学家, 有助于科学数据在不同学科领域、不同技术层面和不同国家地区之间进行共享。RDA 寻找有意愿的国家和地区组织承办会议, 2013 年 3 月 18—20 日, RDA 在瑞典哥德堡召开了第一次全会, 提出了无障碍开放获取研究数据的目标; 第二次全会于 2013 年 9 月 16—18 日在美国华盛顿召开, 会议的目标是在研究数据的共

享方面建立全球合作伙伴关系;2014年3月26—28日在爱尔兰的都柏林举行第三次全会,主题为“数据共享社区:发挥自己的作用”(The Data Sharing Community: Playing YOUR Part),通过主题演讲、展板海报、网络、工作组与兴趣组专题讨论等各种形式,研讨各个学科、数据生命周期各阶段的科学数据问题;2014年9月22—24日,RDA第四次全会在阿姆斯特丹召开,主题为“没有障碍的研究数据共享”(Research Data Sharing without Barrier);2015年3月8—11日,RDA第五次全会在美国加州圣地亚哥召开,认为RDA成立两年来作为一个全球性的组织以新的合作方式促进了数据共享与交换开放环境的形成;2015年9月23—25日,RDA第六次全会在巴黎召开;第七次全会于2016年2月29日至3月3日在日本东京召开,主题为“开放科学时代的数据共享工作”(Making Data Sharing Work in the Era of Open Science);第八次全会将于2016年9月11—16日在美国科罗拉多州丹佛市召开;第九次全会将于2017年4月3—7日在西班牙巴塞罗那召开。

5.2 成立工作组、兴趣组、协调组

各种工作组(Working Groups)、兴趣组(Interest Groups)和协调组(Coordination Groups)在实现RDA职能和目标方面具有重要的作用,研究人员可以在RDA网站上通过注册参与小组交流。截至2016年1月,RDA已经成立21个工作组、44个兴趣组和5个协调组。

(1)工作组主要围绕基础性数据工作建立,职能是解决数据管理中标准、规范、政策、安全之类的问题,如:数据引用工作组(Data Citation WG)、元数据目录工作组(Metadata Standards Directory WG)等。RDA通常与其他机构联合成立工作组,如与美国国家标准化组织(The National Information Standards Organization,简称NISO)建立研究数据集隐私问题工作组(RDA/NISO Privacy Implications of Research Data Sets WG)等。此外,RDA还针对特定学科和主题的数据问题建

立工作组,如生命科学领域数据政策、标准与数据库关联工作组(BioSharing Registry: Connecting Data Policies, Standards & Databases in Life Sciences WG)等。

(2)兴趣组主要将数据工作与专业学科结合起来,解决相关领域中的数据管理问题,如农业数据兴趣组(Agriculture Data IG)、历史和民族志数字实践兴趣组(Digital Practices in History and Ethnography IG)等。兴趣组将数据管理的一般性问题作为研讨的主题,如数据伦理与社会问题兴趣组(Ethics and Social Aspects of Data IG)、数字保存电子基础设施兴趣组(Preservation e-Infrastructure IG)等。RDA还与其他机构联合成立兴趣组,研讨共同感兴趣的数据管理问题,如与CODATA成立法律互操作兴趣组(RDA/CODATA Legal Interoperability IG)等。

(3)协调组是为解决RDA的运行和管理问题而建立的,如RDA组织咨询委员会(RDA Organisational Advisory Board)等;也负责协调RDA业务会议、学术研讨会等的召开,如工作组/兴趣组合作会议(WG/IG Collaboration Meeting)、数字人文研讨会(Digital Humanities Workshop)等。

6 国际图联:探讨图书馆领域科学数据管理实践

国际图联(IFLA)是代表图书馆和信息服务机构以及用户利益的最重要的国际组织。

6.1 研讨科学数据管理的理论与实践

IFLA大会已经涉及科学数据管理方面的主题。2011年8月,第77届国际图联大会保存、保护与信息技术教育培训部分的主题中包含了数字监护教育,并开展了专题研讨^[48]。2012年8月,第78届国际图联大会将科技图书馆分委会的主题定为“图书馆在数据监护、获取和保存中的角色:一个国际的视角”(The Role of Libraries in Data Curation, Access and Preservation: an International Perspective)。会议认为,在科研发展过程中,数据的监护、存取是不可或缺的,会上研讨的

内容包括:开展国家层面的数据监护计划所面临的主要问题;在构建数据监护知识库中,图书馆扮演的角色;参与数据监护的图书馆员的培训;图书馆数据监护可持续发展的财务模式^[49]。此外,会议还发布了一系列研究成果,如《德国图书馆在研究数据监护和保存中的角色:一项调查结果》《大学图书馆和数据研究服务:准备与态度》《加拿大的国家研究数据倡议》《来自图书馆的领导:北安普顿大学的数据管理计划》《图书馆作为电子数据看护服务的经纪人:来自于德国研究基金项目图书馆数字保存的实践见解》《支持数据交换中的图书馆作用》等研究报告或论文^[50]。这些研究成果主要围绕图书馆在科学数据管理领域的实践展开,探讨了图书馆的角色定位、态度见解、实施计划、功能作用等。

6.2 推进图书馆科研数据服务相关研究

IFLA 认为,图书馆可以通过图书馆学的基本法则则应对各种挑战,在提供科研数据管理相关服务方面,图书馆员可以通过开展数据管理计划和元数据咨询、数据发布和保存、数据素养教育等服务,为科研人员提供贯穿于数据生命周期的帮助。《国际图联杂志》(IFLA Journal)将在 2016 年 10 月围绕科研数据服务(Research Data Services)主题出版专刊,目前正在征集有关图书馆科研数据服务最新理论研究、实践经验方面的论文,分主题主要包括:数据素养、数据仓储与馆藏管理、数字数据保存、数据管理计划、元数据与数据分类、数据引用与标识符、数据政策与许可、图书馆数据服务培训等^[51]。

7 小结与启示

近十五年来,上述国际组织开展了一系列与科学数据相关的实践活动,在世界范围内推进了科学数据的管理、共享与利用。这些国际组织中,UN、EU、OECD 属于综合性国际组织,它们主要是制定科学数据管理领域宏观性、原则性的战略、倡议与计划;而数据领域专业性的国际组织 CODATA、RDA 等则是从具体操作层面推动科学数

据管理实践的开展;IFLA 作为国际图书情报界最大的组织,则探索把科学数据管理与图书情报实践相结合,以拓展信息服务边界、提升信息服务水平。

进入新世纪以来,我国已经意识到科学数据的重要价值,早在 2001 年科技部就启动了“国家科学数据共享工程”项目。特别是近年来,随着“大数据”的兴起,我国政府积极应对,推出了一系列有关“大数据”的战略与计划,其中也涉及科学数据领域。国际组织在科学数据管理领域的实践活动对我国推进相关工作具有重要的参考意义,本文认为可以在三个层面借鉴相关的经验和研究成果。

(1)从宏观层面来说,我国要制定与科学数据管理相关的政策、法规以及战略规划等,应广泛吸收国际组织关于科学数据管理与共享的基本原则和精神。例如,《温哥华宣言》《全球科学信息共有倡议》《布加勒斯特宣言》《公共基金资助研究数据存取宣言》《首尔宣言》等文献确立的基本原则和精神,应在结合我国实际情况的基础上得到借鉴和体现。

(2)从中观层面来说,我国科学数据领域的管理和指导机构,可以借鉴国际组织在科学数据管理方面的组织模式和运行机制。例如,在科技部“国家科学数据共享工程”和中科院“科学数据云”建设发展过程中,可以借鉴和参考 CODATA、RDA 等采用的以组织咨询小组和技术咨询小组为指导,以任务组、工作组和兴趣组(工作周期一般为 12—18 个月)为推进器的组织结构,形成以问题为导向的工作模式。这种组织结构和工作模式既可以保障数据管理机构高效、科学的运作,便于过程监督和目标考核,也能快速应对科学研究环境的变化。

(3)从微观层面来说,应鼓励我国的研究机构、科研团队以及科学家等积极参与国际间科学数据管理与共享的交流与合作。CODATA、RDA 皆以促进科学数据全球共享与利用为使命,特别是 RDA 的开放性,降低了机构组织与科学家参与合作的门槛,提高了各参与方之间的交流效率,有

助于全球科学数据共享理念的实现。因此,国内的科学数据中心等机构应该增强开放性,广泛吸引科研机构、科学家等以加盟的方式开展工作,依靠合力推进数据的管理与共享,以适合“数据密集型科研”环境和“大数据”时代的要求。

参考文献

- 1 国务院关于印发“十二五”国家自主创新能力建设规划的通知 [EB/OL]. [2015-12-18]. http://www.gov.cn/zwggk/2013-05/29/content_2414100.htm.
- 2 国务院关于印发促进大数据发展行动纲要的通知 [EB/OL]. [2015-12-18]. http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm.
- 3 International Declaration on Human Genetic Data [EB/OL]. [2015-12-18]. <http://www.unesco.org/new/en/social-and-human-sciences/themes/bioethics/human-genetic-data>.
- 4 Setting the Digital Agenda for the Memory of the World [EB/OL]. [2015-12-18]. http://www.unesco.org/new/en/media-services/single-view/news/setting_the_digital_agenda_for_the_memory_of_the_world.
- 5 Final Report of the Tunis Phase of the WSIS [R/OL]. [2015-12-19]. <http://www.itu.int/wsis/docs2/tunis/off/9rev1.doc>.
- 6 WSIS Forum 2009 Brochure [EB/OL]. [2015-12-19]. http://www.itu.int/wsis/implementation/2009/forum/geneva/docs/WSIS_Forum_2009_Brochure.pdf.
- 7 “Open Access: Progress since WSIS 2010” [EB/OL]. [2015-12-22]. <http://www.unesco.org/new/en/communication-and-information/flagship-project-activities/unesco-and-wsis/implementation-and-follow-up/wsis-forum/wsis-forum-2011/e-science-open-access-to-scientific-knowledge>.
- 8 The Global Information Commons for Science Initiative [EB/OL]. [2015-12-25]. <http://www.codata.org/task-groups/global-information-commons-for-science-initiative>.
- 9 中科院成为全球 UN GAID e-SDDC 领衔单位 [EB/OL]. [2015-12-24]. http://www.edu.cn/ke_ji_chan_ye_1086/20071205/t20071205_268829.shtml.
- 10 UNdata [EB/OL]. [2015-12-29]. <http://data.un.org>.
- 11 UNdata wins 2010 UN21 Award [EB/OL]. [2015-12-29]. <http://unstats.un.org/unsd/wsd/News1.aspx>.
- 12 Agricultural Information Management Standards (AIMS) [EB/OL]. [2015-12-28]. <http://aims.fao.org>.
- 13 Rice genetics placed in global data pool to help develop sustainable crops-UN agency [EB/OL]. [2015-12-29]. <http://www.un.org/News/dh/pdf/english/2015/07102015.pdf>.
- 14 国家测绘地理信息局有关负责人就联合国大会决议《促进可持续发展的全球大地测量参考框架》答记者问 [EB/OL]. [2015-12-29]. <http://www.sbsm.gov.cn/article/gzcy/zxft/201503/20150300022361.shtml>.
- 15 The Bucharest Declaration, Bucharest Pan-European Conference in Preparation of the World Summit on the Information Society: Towards an Information Society: Principles, Strategy and Priorities for Action [R/OL]. [2016-01-05]. <http://www.unecce.org/fileadmin/DAM/env/pp/documents/bucharest.declaration.doc>.
- 16 e-IRG Report on Data Management [R/OL]. [2016-01-05]. http://ec.europa.eu/research/infrastructures/pdf/esfri/publications/esfri_e_irk_report_data_management_december_2009_en.pdf.
- 17 e-Infrastructures in support of the Digital Agenda [R/OL]. [2016-01-05]. http://e-irg.eu/documents/10920/239416/digital_agenda_

- a5_final.pdf.
- 18 Riding the Wave: How Europe can gain from the rising tide of scientific data [R/OL]. [2016-01-05]. http://ec.europa.eu/information_society/newsroom/cf/document.cfm?action=display&doc_id=707.
- 19 Global Research Data Infrastructures: Towards a 10-year vision for global research data infrastructures [EB/OL]. [2016-01-09]. <http://www.grdi2020.eu/Repository/FileScaricati/6bdcd07fb-b21d-4b90-81d4-d909fdb96b87.pdf>.
- 20 HORIZON 2020 The EU Framework Programme for Research and Innovation [EB/OL]. [2016-01-20]. <http://ec.europa.eu/programmes/horizon2020/en>.
- 21 Research Data e-Infrastructures: Framework for Action in H2020 [R/OL]. [2016-01-19]. https://ec.europa.eu/digital-agenda/sites/digital-agenda/files/data_einfra_h2020_fiches_on-line_consult.pdf.
- 22 LIBER/COAR/OpenAIRE joint response to EC Research Data e-Infrastructures Framework [EB/OL]. [2016-01-20]. <http://libereurope.eu/blog/2013/03/28/liber-coar-openaire-joint-response-to-ec-research-data-e-infrastructures-framework>.
- 23 Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 [EB/OL]. [2016-01-28]. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.
- 24 Declaration on Access to Research Data from Public Funding [EB/OL]. [2016-01-15]. <http://archiv.twoday.net/stories/133859>.
- 25 OECD Principles and Guidelines for Access to Research Data from Public Funding [R/OL]. [2016-01-15]. <http://www.oecd.org/science/sci-tech/38500813.pdf>.
- 26 The Seoul Declaration for the Future of the Internet Economy [R/OL]. [2016-01-15]. <http://www.oecd.org/internet/consumer/40839436.pdf>.
- 27 OECD Science, Technology and Industry Outlook 2012 [EB/OL]. [2016-01-19]. http://www.oecd-ilibrary.org/science-and-technology/oecd-science-technology-and-industry-outlook-2012_sti_outlook-2012-en.
- 28 Exploring Data-Driven Innovation as a New Source of Growth [EB/OL]. [2016-01-20]. http://www.oecd-ilibrary.org/science-and-technology/oecd-digital-economy-papers_20716826.
- 29 Making Open Science a Reality [EB/OL]. [2016-01-19]. <http://dx.doi.org/10.1787/5jrs2f963zs1-en>.
- 30 Data-Driven Innovation Big Data for Growth and Well-Being [EB/OL]. [2016-01-09]. http://www.oecd-ilibrary.org/science-and-technology/data-driven-innovation_9789264229358-en.
- 31 International Council for Science: Committee on Data for Science and Technology [EB/OL]. [2015-12-26]. <http://www.codata.org>.
- 32 CODATA Strategic Plan 2006-2012 [EB/OL]. [2015-12-26]. <http://www.codata.info/archives/2006/CODATAstrategicPlan.pdf>.
- 33 CODATA TASK GROUPS 2013-2014 [EB/OL]. [2015-12-26]. http://www.codata.info/resources/newsletters/Newsletter_SpecialIssue_TG.pdf.
- 34 CODATA Strategic Plan 2013-2018 [EB/OL]. [2015-12-26]. <http://www.codata.org/about-codata/codata-strategic-plan-2013-2018>.
- 35 Conference [EB/OL]. [2015-12-30]. <http://www.codata.org/events/conferences>.
- 36 Workshops [EB/OL]. [2015-12-30]. <http://www.codata.org/events/workshops>.
- 37 International Data Week [EB/OL]. [2015-12-30]. <http://www.internationaldataweek.org>.

- 38 Data Science Journal [EB/OL]. [2015 - 12 - 31]. <http://www.codata.org/publications/data-science-journal>.
- 39 刘红, 胡新和. 数据哲学构建的初步探析 [J]. 哲学动态, 2012 (12).
- 40 Data-Intensive Science and Discovery CODATA 45 Years On Summary Report [EB/OL]. [2015 - 12 - 31]. <http://english.ceode.cas.cn/ps/201203/P020120321374100156825.pdf>.
- 41 Preservation of and Access to Scientific and Technical Data in Developing Countries [EB/OL]. [2015 - 12 - 31]. <http://www.codata.info/taskgroups/TGpreservation/index.html>.
- 42 International Training Workshop for Developing Countries on Scientific Data Management and Sustainable Development [EB/OL]. [2015 - 12 - 31]. <http://www.codata.info/archives/2012/index.html>.
- 43 International Training Workshop for Developing Countries on Scientific Data Management and Sustainable Development [EB/OL]. [2015 - 12 - 31]. <http://www.codata.org/Various%20events/CODATA%20Training%20Workshop.pdf>.
- 44 Fundamental Physical Constants [EB/OL]. [2015 - 12 - 31]. <http://www.codata.info/taskgroups/TGfundconst/index.html>.
- 45 CODATA/VAMAS Joint Working Group on the Description of Nanomaterials [EB/OL]. [2015 - 21 - 31]. <http://www.codata.org/nanomaterials>.
- 46 CODATA KEY VALUES FOR THERMODYNAMICS [EB/OL]. [2015 - 21 - 31]. <http://www.codata.org/resources/databases/key1.html>.
- 47 The Research Data Alliance [EB/OL]. [2016 - 01 - 09]. <https://rd-alliance.org>.
- 48 Information Technology Section events and sessions Puerto Rico 2011 [EB/OL]. [2016 - 01 - 21]. <http://www.ifla.org/news/information-technology-section-events-and-sessions-puerto-rico-2011>.
- 49 The Role of Libraries in Data Curation, Access, and Preservation; an International Perspective [EB/OL]. [2016 - 01 - 21]. <http://conference.ifla.org/ifla78/calls-for-papers/the-role-of-libraries-in-data-curation-access-and-preservation-international>.
- 50 The role of libraries in data curation, access and preservation; an international perspective [EB/OL]. [2016 - 01 - 24]. <http://conference.ifla.org/past-wlic/2012/session-116.htm>.
- 51 Call for Papers; IFLA Journal special issue on Research Data Services [EB/OL]. [2016 - 01 - 28]. <http://www.ifla.org/node/9909>.

(黄如花 教授 武汉大学信息管理学院副院长
武汉大学信息资源研究中心, 周志峰 馆员 武
汉大学信息管理学院图书馆学专业 2011 级博士
研究生 温州大学图书馆)

收稿日期: 2016 - 02 - 26