

国外科研数据基础设施研究及实践的调研与分析

王 敬^{1,3} 王彦兵^{2,3} (1 中国科学院兰州文献情报中心 兰州 730000;
2 中国科学院文献情报中心 北京 1000190;
3 中国科学院大学 北京 100049)

摘要 文章在广泛调研国外科研数据基础设施(RDI)的基础上,介绍了 RDI 的定义、发展历程,分析比较了三个具有代表性的 RDI 框架模型以及 RDI 当前的实践状况,同时也讨论了构建 RDI 面临的主要挑战与应对策略。

关键词 科研数据基础设施 数据共享 合作研究

Investigation and Analysis of Foreign Research Data Infrastructure Research and Practice
Wang Jing^{1,3} Wang Yanbing^{2,3} (1The Lanzhou Branch of National Science Library, CAS, Lanzhou, 730000;
2National Science Library, CAS, Beijing, 100190;
3University of Chinese Academy of Sciences, Beijing, 100049)

Abstract Based on comprehensive investigation of current Research Data Infrastructures overseas, this paper examined the definition, the development course and projects of RDI. Three informative frameworks for RDI were analyzed and compared before we presented current practices of RDI. Main challenges for RDI constructing and the corresponding measures were also discussed.

Keywords Research Data Infrastructures, data sharing, collaborative research

1 引言

当前科学研究面临的重大挑战之一是如何应对日益增长的数据洪流,数据洪流改变了科学研究的工作方式,数据密集型科学快速兴起,其被视为实验、理论研究和计算机模拟之外的科学研究第四范式^[1]。数据密集型科学颠覆了传统的科学研究模式,科学研究的进步越来越依赖于如何有效的处理数据。在此背景下,科研数据基础设施(Research Data Infrastructures,RDI)应运而生,RDI 用于处理科学研究过程不断累积的数据与知识,优化不同学科之间的数据流动,在减少资源重复性浪费的同时促进多(交叉)学科的发展,并且能够把科研数据与出版文献整合在一起,最终实现对不断增长科研数据集的有效开发与利用。

随着数据密集型科学的快速兴起,国外政府部门、

科研院所、基金会等科研资助机构逐渐从构建国家级数字保存系统转变为构建世界级的科研数据基础设施,欧盟、澳大利亚以及美国已经把 RDI 的建设与发展提高到了国家战略层次。欧盟委员会“第七框架计划”于 2010 年 2 月资助了 GRDI2020 项目^[2],用以指导全球 RDI 生态系统的建设。2012 年 8 月,澳大利亚政府专门成立了“科研数据基础设施委员会”(RDIC),评估澳大利亚科研数据发展的前景,并为当前及未来 RDI 的投资优化提供建议。美国国家科学基金会(NSF)也正着力构建一个世界一流的,支持高性能计算、数据分析与可视化以及虚拟工作间的科研基础设施。

本文主要从 RDI 的定义、RDI 发展历程、RDI 框架模型、RDI 的应用与实践以及构建 RDI 面临的主要挑战与应对策略的角度出发,对国外科研数据基础设施的发展进行系统的梳理,以期为我国科研数据基础设

施的研究提供策略参考。

2 科研数据基础设施概述

2.1 科研数据基础设施定义

科研数据基础设施的英文表示主要有“Research Data Infrastructures”、“Scientific Data Infrastructures”、“Data Infrastructure”、“Research Infrastructure”等。

维基百科对“数据基础设施”的定义^[9]：是一个可以促进数据分享与消费的电子化基础设施。通常来讲，基础设施是社会经济得以有效运转所需要的基本服务与工具，而数据基础设施则是数据经济不可缺少的基础结构。

澳大利亚 RDI 委员会将其定义为^[6]：通过数据生成、操作以及获取而为科研服务的一系列设施、设备或者工具，它包含数据本身，并依赖成熟的技术与研究人员来建立、实现、操作并使用它。

GRDI2020 项目专家组将 RDI 定义为^[9]：一个以数字化科研数据为中心，包含服务与工具的管理型网络环境，支持：(1)整个研究周期；(2)不同学科之间研究数据的有效流动；(3)通过连接来自不同学科的数据集来创建开放关联数据空间；(4)管理科学工作流；(5)研究数据与文献之间的互操作；(6)一个集成的科学政策框架。

从以上定义可以看出 RDI 是一个包含软件与硬件，支持科研数据的分享与重用，促进全球科研合作，加速科学创新的数据生态系统，是当今数字化科研得以高效运转而不可或缺的支撑性基础设施。

2.2 科研数据基础设施的发展历程

当前，国外对 RDI 的研究主要集中在欧盟、澳大利亚以及美国等地区与国家，其中以欧盟对 RDI 关注的时间最早、研究的最为深入并且规划实施的项目也最多，如表 1 所示。早在 2002 年，欧盟就开始关注科研基础设施的发展，成立了欧洲科研基础设施战略论坛^[6] (the European Strategy Forum for Research Infrastructures, ESFRI)，旨在为欧洲科研基础设施政策的制定提供统一的战略指导，推进欧洲科研一体化，加强国际合作。2016 年，在其首次发布科研基础设施战略路线图的 10 年之后，ESFRI 发布了第四次战略报告《科研基础设施战略报告——路线图 2016》^[7]。报告主要为未来 10~20 年科研基础设施的发展提供战略指导，促进科研基础设施的实现并根据需要适时的更新路线图。报告展示了 15 个正在进行以及新开展的 6 个科研基础设施项目，涉及能源、环境、健康 & 食品、物理科学以及社会文化等领域，ESFRI 将会在未来对这 21 个项目的顺利实施提供必要的战略指导与支持。此外，报告还列举了 29 个已成功实施的科研基础设施项目，并将其作为科研基础成功建设的标杆。

ESFRI 是欧盟开展 RDI 研究的基础，在此期间，欧盟委员会还资助了其他许多项目，如：“科研数据高水平专家组”^[8] (2009)、“绘制欧洲科研基础设施蓝图”

(Mapping of the European Research Infrastructure Landscape, MERIL, 2010)^[9]、GRDI2020 项目^[6] (2010)、“泛欧洲协作数据基础设施”(EUDAT, 2011)^[10]以及“知识交流”(Knowledge-Exchange, KE, 2012)^[11]等。“科研数据高水平专家组”在其发布的报告中探讨了如何从科研数据浪潮中获益，指出构建 RDI 的益处、所面临的挑战，并提出了相关建议^[12]。与之对应，KE 从丹麦、德国、新西兰以及英国四个国家 RDI 的研究现状出发，提供构建 RDI 的可行性行动计划，归纳构建 RDI 的四个驱动因素^[9]。GRDI2020 项目则直接呼吁构建一个可以支持数据共享与科研合作的全球化科研数据基础设施，并于 2011 年、2012 年发布了两个重要报告，分别探讨大数据背景下构建全球化 RDI 的必要性、所面临的主要挑战^[6]以及构建 RDI 过程中所面临的技术与组织问题^[14]，GRDI2020 项目代表了当前 RDI 发展的最新趋势。

紧随其后，澳大利亚与美国也逐渐把 RDI 的规划与建设提到国家战略层面。2006 年澳大利亚政府实施了“协作科研基础设施战略”(the National Collaborative Research Infrastructure Strategy, NCRIS)，并相继发布了三次战略报告，详细论述为促进科学创新而应优先发展的科研基础设施项目^[15-17]。在此基础上，2012 年，澳大利亚政府成立了“科研数据基础设施委员会”(the Research Data Infrastructure Committee, RDIC)，专门负责规划国家 RDI 的发展并为 RDI 的建设与优化提供建议，同时发布澳大利亚科研数据基础设施战略报告^[6]，把 RDI 的发展提到了前所未有的战略高度。美国国家科学基金会(NSF)也于 2007 年开始关注科研基础设施的发展，在其发布的报告中倡导构建一个世界一流的，支持高性能计算、数据分析与可视化以及虚拟工作间的科研基础设施^[18]。2013 年欧盟、澳大利亚与美国共同倡导成立了 RDA (科研数据联盟)^[19]，RDA 致力于减少数据分享与交换障碍，为科研数据相关服务提供建议，RDI 是其核心议题之一。

从以上各个国家科研数据基础设施的发展历程可以发现，各国最先关注的是科研基础设施的发展。随着大数据与数据密集型科学的兴起，数据成为科研活动中的“一等公民”，研究人员开始把关注的焦点转移到科研数据上，探讨如何构建一个以科研数据为中心的基础设施，从而实现科研数据的充分开发与利用。科研数据基础设施来源于科研基础设施的发展，两者一脉相承，但又有差别，科研基础设施一般指的是支持科研活动的设施、资源与活动，它已演变成为科研数据基础设施的一部分。除此之外，科研数据基础设施还包含可以支持科研数据生命周期全部阶段的基础设施。

3 科研数据基础设施框架模型

构建一个支持全球合作的科研数据基础设施是一项艰巨的任务，当前还面临许多技术、组织与系统挑

表1 国外科学数据基础设施发展历程

国家 & 地区	核心研究团体 & 项目	时间	使命 & 愿景	主要成果
欧盟	ESFRI (欧洲科研基础设施战略论坛)	2002年	为欧洲科研基础设施政策的制定提供统一的战略指导,推进欧洲科学研究一体化,加强国际合作	自2006年至2016年共发布四次ESFRI路线图报告 ^[7,20-22] ,是RDI研究的基础
	欧盟委员会科研数据高水平专家组	2009年	探索欧洲如何从科研数据潮流中获益	于2010年向欧盟委员会提交了最终报告,对构建RDI政策具有重要指导意义
	MERIL (绘制欧洲科研基础设施蓝图)	2010年	为科研人员与政策制定者提供欧洲现存科研基础设施的详细目录信息	构建了一个可以查看欧洲各个国家已建设科研基础设施信息的门户
	GRDI2020项目,全球科研数据基础设施10年展望	2010年	构建一个可以支持数据共享与科研合作的全球化科研数据基础设施	分别于2011年、2012年发布了两个重要报告,对RDI的建设影响深远
	EUDAT (泛欧洲协作数据基础设施)	2011年	使任何学科的欧洲科研人员都能在一个可信任的环境中处理数据	为15个欧洲国家提供一系列科研数据服务、专业知识与技术解决方案
	KE (知识交流)	2012年	通过构建一个国际化信息基础设施,实现科研资源开放共享	针对GRDI2020的战略愿景,提出构建全球化RDI的可行方案
澳大利亚	NCRIS (澳大利亚协作科研基础设施战略)	2006年	为澳大利亚科研基础设施的建设提供战略指导	发布协作科研基础设施战略路线图报告
	RDIC (澳大利亚科研数据基础设施委员会)	2012年	向澳大利亚政府提供关于RDI建设与优化的建议	于2012年发布《澳大利亚科研数据基础设施战略报告》
美国	NSF (美国国家科学基金会)	2007年	构建一个可以支持高性能计算、数据分析与可视化以及虚拟工作间的世界一流科研基础设施	于2007年发布报告《支撑21世纪科学发现的基础设施展望》
欧盟、澳大利亚与美国	RDA (科研数据联盟)	2013年	关注基础设施发展,减少数据分享与交换障碍	为科研数据相关服务提供建议

战。但是,我们需要一个战略性、概念性的框架来指导不同机构科研数据基础设施的建设。欧盟委员会科研数据高水平专家组、澳大利亚科研数据基础设施委员会(RDIC)等机构与个人构建了RDI框架模型,下面将对这些框架进行分析比较,探讨当前RDI所要实现的核心功能与目标。

欧盟委员会科研数据高水平专家组认为RDI应能够支持对科研数据的获取、使用、重用与验证,全球范围内不同领域的科研人员可以在共享的数据集上进行合作研究,在数据分享时也能保护数据的完整性与所有权。专家组构建的RDI框架模型如图1所示(框架1)。

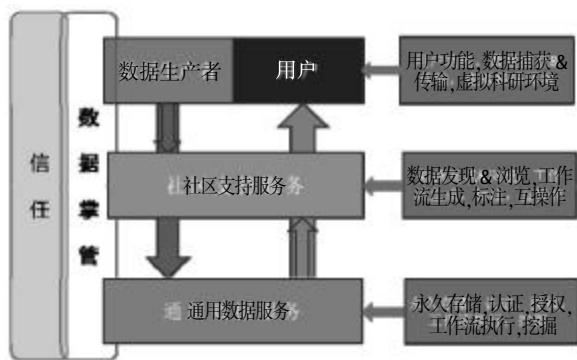


图1 欧盟委员会科学数据高水平专家组 RDI 框架模型^[23]

此框架是一个流动循环系统,它揭示了不同参与

者、数据类型与服务在全球化科研数据基础设施中应如何互相关联。世界各地的数据生产者与用户在虚拟研究环境中收集、捕获与传输数据,并利用社区支持服务发现、使用、标注或解释这些数据。与此同时,利用通用数据服务对数据进行永久存储、认证、授权与挖掘等。数据掌管与信任贯穿整个系统,数据掌管主要处理数据的存储、管理与保存,而信任模块则负责处理数据可信性问题。

澳大利亚 RDIC 认为 RDI 框架必须能够协调国家、领域以及机构等不同层次的投资,避免重复建设,减少浪费,在有限的资源下优先发展重点 RDI,最大限度满足利益相关者的需求。与此同时,框架还需明确 RDI 系统采集科研数据的具体要求,将科研数据作为增值资产进行组织管理,并保证在未来数据驱动、变化的环境下能够使用这些数据。委员会对建成一个成功的科研数据基础设施提出了三点关键要求^[24]:(1)持续、优先的科研数据采集;能够在国家、研究领域以及机构等层次生产与管理科研数据的基础设施。(2)科研人员可以从机构与行业数据中获益的良好环境;清晰、有效的数据治理与获取装置。(3)输出符合国家优先发展战略、更多高质量的科研成果;更好的分享与重用科研数据。

基于这些需求,委员会提出的科研数据基础设施框架如图2所示(框架2)。



图2 澳大利亚科研数据基础设施框架^[24]

该框架以 RDIC 所提出的三点关键性要求为核心,把构建 RDI 的一系列原则与活动划分为“目标”(Objective)、“行动”(Action)与“着力点”(Focus)三大模块。这些约定的原则是政策制定者、投资方、开发商、运营商以及用户建设维持一个高效、完整的 RDI 系统的基础。框架的“行动”模板体现了 RDI 的核心功能:有意、系统的采集数据;组织数据使其可被发现与获取;

以尽可能多的方式不断重复使用数据。

阿姆斯特丹大学的 Demchenko Yuri 研究员根据科研数据生命周期管理模型与现代 e-science 的科研数据管理特性,提出了以共享数据或项目为中心的科研数据基础设施架构模型,如图 3 所示(框架 3)。

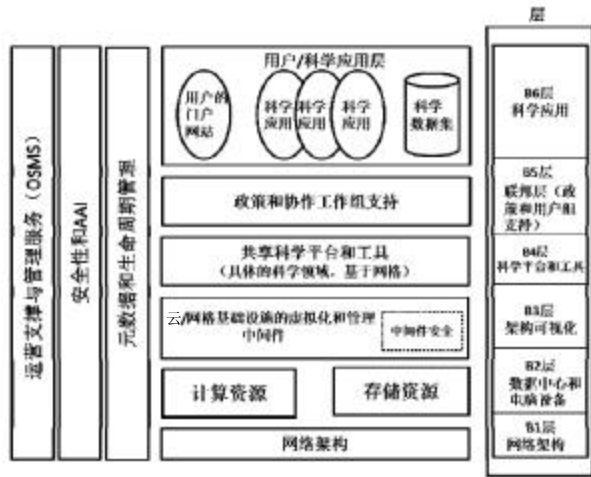


图 3 Demchenko Yuri 科研数据基础设施架构模型^[24]

该框架模型主要包含以下六个层次^[24]:(1)网络基础设施层(B1):基于通用互联网基础设施和专用网络基础设施;(2)数据中心和计算资源/设施(B2层);(3)基础设施虚拟层(B3层):以支持专门的科研平台进行部署和运行的云/网络基础设施服务和中间件为代表;(4)科研平台与基础设施(B4层):专门为不同研究领域提供数据共享的科研平台与基础设施;(5)联合决策层(B5层):包括联合基础设施组件,能够支持决策和协作用户;(6)科研应用和用户门户网站/客户端(B6层)。

另外,模型还包含三个交叉层面:运营支撑与管理服务、安全控制层以及元数据与生命周期管理层,这三层贯穿到整个科研数据生命周期。

对比分析以上三个 RDI 框架模型,三者的目标功能是一致的,均支持科研数据生命周期的全部阶段:数据采集、存储、治理、分析、发现、重用以及共享(出版)等,实现了对科研数据的“一站式”处理与操作,提高了科研效率;允许不同的研究团队围绕相同的数据集进行研究,构建可靠的(虚拟)科研和协作环境,安全的存储中间结果,共享科研成果。此外,框架 1 与框架 3 还都重点考虑了数据的安全性。RDI 具有远程分布的特性,用户可以在不同的国家或地区获取数据,RDI 必须具有安全可靠的数据存储与处理环境,能够保障数据的完整性、机密性与可计量性,不被非授权者获取,同时还需要执行并完善数据的隐私保护政策。三个 RDI 框架所不同的是:框架 1 是以“数据服务”为核心的系统,数据生产者与用户从数据服务中获取价值;框架 2 则以“构建要求”为核心内容,强调构建国家级 RDI 应该遵守的原则与要求;框架 3 则强调利用现代技术与

最佳实践来构建以共享数据或项目为中心的 RDI。前两者是国家战略层次的 RDI 建设指导框架,宏观规划了 RDI 建设的目标、原则与功能,框架 3 则更加具体。就适用范围而言,前两个 RDI 框架更具广泛的指导性与适用性,而框架 3 可以看作是在以上两个模型指导下所构建的更加具体的 RDI 实施框架。这三个框架模型为我们建设 RDI 提供了前瞻性的指导,值得我们思考并实践。

4 科研数据基础设施的应用与实践

当前 RDI 建设呈现出种类繁多、碎片化、变化不定的特点,而且不同国家不同学科的 RDI 组织管理方式也大相径庭,导致这种状况的部分原因在于科研数据集来源与存储的差别。科研数据的来源与存储一般分为四种情况(见表 2)。在项目完成后,由科研个体或者小的研究团体产生的数据集一般由他们自己保存或者存储在他们的机构知识库中。当这些数据集与期刊文章关联时,它们大多被期刊出版商或者机构转移到知识库或者数据档案馆中。来自大的研究团体的数据集则通常存储在数据档案馆(知识库)或者直接出版为数据期刊;而来自研究设施或政府研究机构的数据集,通常存储在它们自己的数据中心,以备将来重复使用。

表 2 科研数据集的来源与存储^[24]

项目进行中 科研数据集类型	项目完成后	
	存储位置	主要用途
来自科研个体或小的研究团体的数据集	短期存储在研究团体或者机构知识库中	验证
	与期刊文章关联的重复数据集,有时由出版商保存,有时存储在机构知识库中	验证
来自大的研究团体或者合作团体的数据集	存储在数据档案馆或知识库中	重用
来自研究设施的数据集,通常有它们自己的数据中心	出版为数据期刊	重用
来自政府研究机构的数据集,通常有它们自己的数据中心	大部分存储在数据中心	重用

根据其规模大小,可以大致把当前建设的 RDI 分为以下两种类型:

(1)机构数据基础设施。有些研究机构构建的是虚拟科研环境,研究人员可以在此环境下共同合作并生产、收集以及处理科研数据;还有一些机构构建的是可以存储数据并可与他人分享数据的机构数据知识库。虚拟科研环境与数据知识库都是机构数据基础设施的一部分,它们还处于发展初期,只存在于很少一部分研究机构,而且非常分散。

(2)学科或者国家(国际)数据基础设施。以学科或者国家(国际)的类别所构建的数据中心(数据档案馆)类型多样。学科数据中心一般由科学社区根据它们的需求而成立。例如:地球与环境科学的世界数据中心系统(The World Data Centre System)^[27],它包含分布在全球 12 个国家的 52 个数据中心,另外,还有基因组数据

库以及国际天文学虚拟观测联盟等。欧盟正逐步建设涉及多个学科、国际性的数据基础设施,如上文提到的MERIL与GRDI项目,其中MERIL项目已初步完成,其构建了一个可以公开获取欧洲范围内各个学科领域研究基础设施信息的数据库。表3列出了部分欧洲国家已建成、具有代表性的“学科或者国家(国际)数据基础设施”。

表3 欧洲部分国家数据基础设施^[3]

丹麦	德国	荷兰	英国
丹麦数据档案馆	地球与环境科学世界数据中心	数据存档与网络服务(DNAS)	考古学数据服务
欧洲环境机构数据服务	德国地学研究中心 地震数据档案馆	土壤世界数据中 心	生物记录中心
地磁学世界数据中心	气候世界数据中心	马普学会语言档 案馆	英国大气数据中心
	大气遥感技术世界 数据中心	欧洲海洋环境数 据目录(EDMED)	英国海洋数据中心
	社会科学 GESIS 数 据档案馆	射电望远镜 (LOFAR)	化学数据库服务
	德国卫星数据档案 馆	荷兰气象局 (气象学)	结晶学数据服务
	DNA 银行网络		环境信息数据中心
			欧洲生物信息学协 会
			国家地球科学数据 中心
			冰川冻土世界数据 中心
			极地数据中心
			英国数据档案馆
			视觉艺术数据服务

从表3可以看出,当前各国科研数据基础设施的建设大多是基于学科领域,其中以英国建设的数据中心(档案馆)数量最多,涉及的学科也最广泛(限于篇幅,只列出了一部分)。当前还没有建成可以支持在世界范围内跨学科合作与共享的全球化科研数据基础设施。

5 构建科研数据基础设施面临的主要挑战与应对策略

虽然有些国家已经建成了部分科研数据基础设施,但是这些RDI是分散的、孤立的并且大多是基于某个学科领域的,远没有达到全球共享数据、合作研究的最终目标。构建这样一个支持跨(多)学科、共享数据与合作研究的全球化科研数据基础设施面临着诸多挑战。

(1)数据共享。数据共享是RDI建设的核心,如何激励科研人员分享他们的科研数据是构建RDI所面临的巨大挑战。一项针对研究人员的国际调查显示,只有25%的研究者分享了他们的数据,法律问题与对共享数据可能的滥用是阻碍数据分享的两大障碍^[28]。虽然这种状况正逐渐好转,但是数据分享仍旧只局限于小部分研究者,而且分享的数据集也只是一小部分,数据分享还没有成为学术文化不可或缺的一部分。要解决这

种困境,一种可能的做法是可以把RDI培育成为一个可以代表科研人员成果的“数据收录”系统。研究人员的数据被RDI“收录”后,将有助于促进研究人员的职业发展,同时给其带来学术声望。

(2)数据互操作。一方面,当前已建成的RDI大多是基于某个学科领域或机构,未来的RDI建设不可能把现有RDI全部推倒重建,最佳方案是把当前已建成的RDI连接起来,形成一个开放、动态、共享的RDI网络;另一方面,RDI的核心功能是数据共享,不管是什么类型的RDI都必须考虑数据互操作的问题。因此,RDI之间的数据互操作将会是RDI建设过程中必须要考虑的重要问题。若要实现全球范围内RDI的互操作,统一的软件技术标准、一致的数据模型与网络协议等将会是RDI建设者需要重点考虑的因素。

(3)数据信任。由上文欧盟RDI框架可知,数据信任问题贯穿RDI框架模型的每个层次,它包含数据的认证与授权。RDI为科研人员提供的是一个可以存储、获取并分享数据的分布式、虚拟的科研环境,用户可以在不同的国家获取数据,在此环境下,RDI如何确保数据的真实性与可靠性?面对此问题,一方面需要数据分享者提供有关数据的详细来源信息以及数据的应用场景,以辅助数据使用者初步判定数据的有效性;另一方面,RDI需要召集相关的领域专家对其进行初级认证并鼓励全球范围内的数据使用者对数据进行信用评级,级别越高说明数据越可靠。

(4)数据保存。众所周知,数据存储介质在不断演化,从之前的电磁线圈、磁带到今天普遍使用的硬盘,以及方兴未艾的云存储,存储介质的不断变化、升级将极有可能导致存储信息的丢失。RDI如何能保证在未来很长一段时间内将所存储的数据保存完好,并像现在一样能被大多数用户而不是特别的专家较容易的存取使用?数据长期保存问题的解决将依赖于未来成熟、稳定、高效的数据迁移工具,在保障数据完整性的情况下,把存储的数据迁移至新型的存储介质上。

(5)其他挑战。如:未来RDI必须支持数据分析、数据可视化与数据挖掘等活动,因此,RDI就需要开发出众多数据处理工具;虚拟研究环境对于数据密集型科学至关重要,RDI必须提供可以构建、支持和维护虚拟研究环境的架构与管理工具。

构建RDI所面临的以上挑战在普通数据存储系统中也有可能面临,但是与普通数据存储系统不同的是,RDI是一个全球化的数据基础设施,它的目标不是存储数据,而是在世界范围内共享与使用数据,它需要处于不同文化背景、不同经济发展水平的国家参与其中。因此,它所面对挑战将会更加综合、复杂,处理起来也更加棘手。

6 结语

建设科研数据基础设施的目标不是构建一个世界范围内的超级数据存储中心,而是构建一个可以分享数据、促进研究者合作与交流,从而加速创新,为世界发展提供动力的数据基础设施。虽然世界各地建成了一些科研数据基础设施,但是这些 RDI 是分散的、孤立的,远没有达到全球共享数据、合作研究的最终目标,这就需要世界各国的研究机构携手合作,共建共享。图书馆作为信息资源和信息服务的提供者同样在 RDI 建设过程中扮演着重要角色,图书馆特别是研究型图书馆,是一个机构乃至一个国家的信息中心,它负责为用户提供信息资源与服务,参与国家信息发展战略的制定。因此,一方面,图书馆可以利用自身的优势为 RDI 的建设提供战略咨询并协助制定 RDI 国家发展战略;另一方面,图书馆作为 RDI 建设的倡导者与协调者,它可以积极推动将分散在各个区域的数据中心连接起来,形成一个巨大的 RDI 网络,最后实现与国际 RDI 的连接与共享。

参考文献

- [1] Hey T, Tansley S, Tolle K. The fourth paradigm: data-intensive scientific discovery[J]. General Collection, 2009, 317(8):1.
- [2] GRDI2020[EB/OL]. [2015-09-15]. <http://www.grdi2020.eu/Default.aspx>.
- [3] Data Infrastructure [EB/OL]. [2015-09-20]. https://en.wikipedia.org/wiki/Data_infrastructure.
- [4] The Australian Research Data Infrastructure Strategy[EB/OL]. [2015-05-20]. https://docs.education.gov.au/system/files/doc/other/the_australian_research_data_infrastructure_strategy.pdf.
- [5] Global Research Data Infrastructures: The Big Data Challenges [EB/OL]. [2015-01-20]. <http://www.grdi2020.eu/Repository/FileScaricati/fc14b1f7-b8a3-41f8-9e1e-fd803d28ba76.pdf>.
- [6] ESFRI[EB/OL]. [2015-09-20]. http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri.
- [7] Strategy Report on Research Infrastructure Roadmap 2016[EB/OL]. [2016-04-25]. http://www.esfri.eu/sites/default/files/20160309_ROADMAP_browsabte.pdf.
- [8] High Level-Expert Group on Scientific Data[EB/OL]. [2016-04-20]. <http://www.grdi2020.eu/Pages/Unlock.aspx>.
- [9] MERIT [EB/OL]. [2016-04-23]. <http://portal.merit.eu/conversesf/static/about>.
- [10] EUDAT[EB/OL]. [2015-09-20]. <http://eudat.eu/>.
- [11] KE[EB/OL]. [2015-09-10]. <http://www.knowledge-exchange.info/about-us>.
- [12] Riding the Wave-How Europe Can Gain from the Rising Tide of Scientific Data[EB/OL]. [2016-03-23]. <http://www.grdi2020.eu/Repository/FileScaricati/c2194260-3ddf-47bd-93e4-68f8912a3564.pdf>.
- [13] A Surfboard for Riding the Wave - Towards a Four Country Action Programme on Research Data [EB/OL]. [2015-04-16]. [ary%20Research%20Data/Surfboard%20for%20Riding%20the%20Wave/KE_Surfboard_Riding_the_Wave_Screen.pdf.](http://www.knowledge-exchange.info/Files/Filer/downloads/Prim

</div>
<div data-bbox=)

- [14] Technological & Organizational Aspects of a Global Research Data Infrastructure -A View from the Experts[EB/OL]. [2016-02-02]. <http://www.grdi2020.eu/Repository/FileScaricati/9a85ca56-c548-47e4-8b0e-86c3534ad21d.pdf>.
- [15] Strategic Roadmap for Australian Research Infrastructure Roadmap 2006 [EB/OL]. [2016-01-25]. https://docs.education.gov.au/system/files/doc/other/national_collaborative_research_infrastructure_strategic_roadmap_2006.pdf.
- [16] Strategic Roadmap for Australian Research Infrastructure Roadmap 2008[EB/OL]. [2016-01-20]. <http://www.industry.gov.au/science/Documents/StrategicRoadmap.pdf>.
- [17] 2011 Strategic Roadmap for Australia Research Infrastructure [EB/OL]. [2015-12-05]. https://docs.education.gov.au/system/files/doc/other/national_collaborative_research_infrastructure_strategic_roadmap_2011.pdf.
- [18] CyberInfrastructure Vision for 21st Century Discovery[EB/OL]. [2016-02-27]. <http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf>.
- [19] RDA[EB/OL]. [2015-09-11]. <https://rd-alliance.org/>.
- [20] Strategy Report on Research Infrastructure Roadmap 2006[EB/OL]. [2016-02-24]. http://ec.europa.eu/research/infrastructures/pdf/esfri/esfri_roadmap/roadmap_2006/esfri_roadmap_2006_en.pdf#view=fit&pagemode=none.
- [21] Strategy Report on Research Infrastructure Roadmap 2008[EB/OL]. [2016-02-22]. http://ec.europa.eu/research/infrastructures/pdf/esfri/esfri_roadmap/roadmap_2008/esfri_roadmap_update_2008.pdf#view=fit&pagemode=none.
- [22] Strategy Report on Research Infrastructures Roadmap 2010[EB/OL]. [2016-03-01]. https://ec.europa.eu/research/infrastructures/pdf/esfri-strategy_report_and_roadmap.pdf.
- [23] Research Data e -Infrastructures:Framework for Action in H2020[EB/OL]. [2016-03-24]. http://cordis.europa.eu/fp7/ict/einfrastructure/docs/framework-for-action-in-h2020_en.pdf.
- [24] Yuri D, Zhiming Z, Paola G, 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, 45(1): 30-35,40.
- [25] Demchenko Y, Zhao Z, Grosso P. Addressing Big Data Challenges for Scientific Data Infrastructure[C]. Proceedings of 2013 International Conference on Collaboration Technologies and Systems, San Diego, CA. IEEE, 2013: 48-55.
- [26] SURFfoundation[EB/OL]. [2015-09-18]. <http://www.surfoundation.nl/>.
- [27] WDC[EB/OL]. [2015-09-29]. <http://www.icsu-wds.org/>.
- [28] Insight into Digital Preservation of Research Output in Europe [EB/OL]. [2015-09-06]. http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf.

[作者简介]王敬,男,1988年生,中国科学院大学、中科院兰州文献情报中心硕士研究生。

王彦兵,男,1985年生,中国科学院大学、中科院文献情报中心硕士研究生。

收稿日期:2016-03-04