

# 国家科研数据服务实践进展及启示

□ 刘晓娟 于佳\* 林夏

**摘要** 基于美国、英国、加拿大、澳大利亚、荷兰、瑞典6个国家科研数据服务的实践,从服务概况、基础设施建设、科研数据的关联、发现、再利用及合作机构等方面总结了各国从国家层面提供科研数据服务的经验。研究表明,国外在国家层面的科研数据服务方面已经取得了一定的进展。相比之下,我国的科研数据服务目前主要集中在基础科学领域,并且主要提供以学科为导向的数据资源集成服务。我国需要在借鉴学习国际成功经验基础上,加大政府的引导和支持力度,宏观规划国家科研数据服务支撑项目,有效整合各机构和各学科领域的科研数据资源和基础设施,深化科研数据服务层次,最终构建具有中国特色的国家科研数据服务体系。

**关键词** 科研数据服务 数据发现 数据关联 数据再利用

**分类号** G203

**DOI** 10.16603/j.issn1002-1027.2016.05.005

## 1 引言

在大数据背景下,科学研究以数据为中心的特征越来越突出,围绕数据展开的科研数据服务也应运而生,如高校图书馆等信息服务机构向科研人员提供科研数据管理计划(DMP)的撰写指导、科研数据组织、归档、分析、长期保存、共享,以及提供相关工具与资源推荐等。此外,在不同服务场景中也涌现出了与科研数据服务相近的概念,如科研数据管理(RDM)、数据监护(Data Curation)、数据归档(Data Archiving)等,它们都是针对科研数据开展的相关服务,但各有侧重,如科研数据管理重点涵盖了数据管理标准制定、元数据创建、语义注释、数据关联、数据发现与共享等活动。相对来说,数据监护与数据归档则更强调科研数据的长期存储、获取与再利用等环节。然而这些服务多针对某个学科领域,或由各机构或联盟分散进行,缺乏统一的政策指导与标准规范,逐渐出现“数据孤岛”的问题。这种现实的情况迫使科研数据管理工作必须进行变革,但仅靠高校或科研机构的单一力量难以快速推进,需要有一个能够将各个利益群体统一协调管理的组织,例如代表国家层面的相关机构。近年来,美国、英国、澳大利亚等国家都已经提出了国家科研数据

服务的理念,通过不同的方式在国家层面开展了科研数据相关工作,致力于整合科研数据及基础设施,为科研人员提供高效高质的科研数据服务。

经调查发现,目前科研数据服务的研究成果多集中于科研院校<sup>[1-3]</sup>、高校图书馆<sup>[4-5]</sup>、学科领域<sup>[6]</sup>、出版机构<sup>[7-8]</sup>及科研资助机构<sup>[9-10]</sup>等层面,而从国家层面开展科研数据服务的研究尚不完善,目前主要包括以下几个方面:(1)探讨开展国家科研数据服务的必要性。利兹·里昂(Liz Lyon)于2007年提出联邦数据存储模型的理念,并指出基金会对所资助的项目实施有效的管理,实现更高层面的科研数据基础设施和服务的整合,对于应对数据洪流有着重要的意义<sup>[11]</sup>。(2)探讨国家科研数据服务主体的职责分工。樊俊豪从整个学术交流体系的角度对科学数据服务中各相关主体的作用进行了探讨,认为政府和基金会作为重要行为主体,可以从国家层面部署科学数据管理的基础设施,制定相应的政策等,最终构建新的数据驱动的经济社会发展范式<sup>[12]</sup>。(3)正在开展的国家科研数据服务项目的经验分享。2011年加拿大国家科研数据峰会的会议报告中分析了澳大利亚、美国、英国、德国等国的科研数据管理开展情况,并提出了加拿大开展国家科

\* 通讯作者:于佳,ORCID:0000-0002-4835-4970,happyujia@163.com。

研数据服务的计划和步骤<sup>[13]</sup>。安德鲁·特雷洛尔(Andrew Treloar)多次在报告中对澳大利亚国家科研数据服务的进展情况进行了分享<sup>[14]</sup>。

虽然许多国家层面的科研数据服务已经投入使用,却少有研究者对这些项目进行综合性的对比与评述。文章参考科研数据联盟(RDA)及欧洲数据基础设施项目(EUDAT)的成员名单,选取了具有代表性和借鉴意义的6个对象加以研究,分别为美国科研数据服务(National Data Service, NDS)、英国科研数据服务(UK data Service, UKDS)、加拿大科研数据服务(Research Data Canada, RDC)、澳大利亚科研数据服务(Australian National Data Service, ANDS)、荷兰科研数据服务(Data Archiving and Networked Service, DANS)及瑞典科研数据服务(Swedish National Data Service, SND),通过对这些典型案例的分析,对我国从国家层面开展科研数据服务提供借鉴和参考。

## 2 国家科研数据服务概况

国家科研数据服务是指充分发挥全国范围内的政府机构、基金会、计算中心、数据仓储库、科研机构、信息服务机构等利益群体的协作效果,在数据存储的基础上,为科研人员提供跨领域的发现、共享与再利用等服务,使得科研数据具备更准确的描述性、更好的发现性、更紧密的关联性、更强的集成

性和更便捷的获取性等特点。

美国、英国、加拿大、澳大利亚、荷兰、瑞典等国家非常重视科研数据所具有的科学价值、经济价值和社会价值,如图1所示,各国逐步开展大量的数据管理实践,并尝试从国家层面提供科研数据服务。

在国家科研数据服务的产生背景方面,各国根据实践情况不同而有所区别。美国和英国的科研资助机构陆续颁布了相应政策,要求研究者提交数据管理计划<sup>[15]</sup>,并且已经成功开展了很多数据服务相关的项目。如2007年美国的数据长期保存项目(Data Conservancy)采用OAI-ORE标准,建立了发布、管理和维护跨学科观测数据的模型架构<sup>[17]</sup>;英国也已经在很多学科领域建设了比较完善的数据中心及仓储库<sup>[18]</sup>。澳大利亚和加拿大的国家科研数据服务主要是作为政府构建科研基础设施和网络的重要组成部分,用于提升国家科研竞争力,如ANDS作为澳大利亚国家合作研究基础设施(NCRI)项目<sup>[16]</sup>的重要组成部分,与高性能计算设施项目(NCI)、科研无缝链接项目(AREN & AAF)等共同建构了面向科研人员的、具有世界级水平的科研基础设施。瑞典与荷兰主要是依托于人文社科、医学等特定领域的科研数据服务实践,逐步完善面向全国层面的科研数据服务,改善科研数据的应用大环境。

### 澳大利亚 ANDS

创建时间:2009年



学科领域:全部学科领域

### 瑞典 SND

创建时间:2008年



学科领域:社会、人文科学及医学领域

### 荷兰 DANS

创建时间:2005年



学科领域:社会、人文科学领域

### 加拿大 RDC

创建时间:2011年



学科领域:全部学科领域

### 英国 UKDS

创建时间:2012年



学科领域:社会、人文科学及经济学领域

### 美国 NDS

创建时间:2014年



学科领域:全部学科领域



图1 国家科研数据服务的典型案例

在服务目标方面,各国都致力于通过全国范围内的科研数据整合,优化国内的数据环境,促进科研创新。其中 ANDS 更将科研数据提升到了国家战略资源的高度进行管理,RDC 则强调要实现从广泛的公共投资的科研项目中获得更大更长远的效益。SND、UKDS 和 DANS 关注支持国家层面的特定学科领域的数据,通过对该学科领域现有的科研数据库的整合,促进标准规范、文档管理的相关方法和工具技术的发展,最终提升科研数据再利用的效率。

在服务对象方面,各国国家科研数据服务对象都普遍包括国内的科研机构及研究人员,同时由于自身的服务定位不同,各国在服务对象的群体上也有些细微的区别。NDS 特别指明服务对象中包括大规模科研项目和科研团队。UKDS 则提到可以为政策分析人士、慈善组织和基金会、企业顾问和数据分析师、独立研究中心等提供相应的科研数据服务。由于 DANS 的服务目标之一是促进荷兰人文社科领域的高质量的研究与教育,因此高校学生是其主要服务对象之一。

### 3 基础设施

国家科研数据服务的有效开展需要有完善的基础设施作为支撑。根据科研数据联盟(RDA)的划分标准,科研数据基础设施可分为技术基础设施和社会基础设施,前者是指实现科研数据获取、存储、处理、发布与共享所需要的一系列相互联系的平台、硬件、软件、工具组成的基本架构以及相关技术标准(如数字标识符、共享的元数据框架等),后者主要指促进不同机构间协作的措施和手段,如公共政策和组织实践、统一的标准、通用的数据获取和保存方法等<sup>[19]</sup>。

虽然近年来才开始有国家科研数据服务的提出和建设,但针对科研数据已经有很多研究成果和实践进展,所以国家科研数据的建设并不是从零开始,而是需要充分利用各机构、各领域现有的基础设施,其中重点是如何将适用于特定机构和领域的基础设施,通过平移、包装、映射等操作增强其通用性,从而适应跨机构和跨领域的科研数据服务需求。因此在国家科研数据服务的建设中,不论是技术基础设施,还是社会基础设施,都存在着从局部适用到全局适用的转换。

在技术基础设施建设中,要充分利用已有的平

台和软件工具,在保留其原有优势的同时克服存在的问题,从而提升其服务的性能。此外为了跨越空间的限制,需要利用通用的元数据框架及收割协议等技术标准,扩大服务范围,实现从机构层面科研数据服务向国家层面的过渡。以澳大利亚 ANDS 为例,在其与莫纳什大学合作的“多媒体科研数据集监护及发布工具”项目中充分利用了图书馆已有的数据仓储库“箭头”(ARROW)的各项功能,同时整合了新的图像管理工具 Picasa,可以实现创建图像数据集、增加描述性元数据、处理原始图像等高级功能,提升了莫纳什图书馆对影像数据集的监管服务。与此同时,数据仓储库中的图像数据集可以通过 RIF-CS 元数据框架、以及 OAI-PMH 等收割协议最终被 ANDS 收割,通过澳大利亚科研数据发现系统(Research Data Australia)从国家层面实现对这些数据集的发现和获取。

社会基础设施主要包括借鉴传统的数字对象处理经验,在政策框架和通用方法等方面达成共识,构建有利于科研数据共享与再利用的大环境。如 ANDS 利用已经在政府部门得到广泛应用的政府信息许可框架(AuaGOAL),通过与科研社区和基金组织合作,将该框架的适用性延伸到澳大利亚所有科研及创新领域,澳大利亚各部门和机构的数据拥有者可以参考 AuaGOAL 选择适当的许可协议,在发布相关信息时可以更好的保护数据的版权、隐私或商业机密。此外,UKDS、DANS 等也纷纷围绕科研数据生命周期推荐了众多“最佳实践”服务案例,尝试建立能被各机构认同和接受的通用标准及方法。

### 4 科研数据的关联

科研数据的关联是指在基础设施的支持下,通过特定策略使资源贡献者在描述科研数据、添加数据之间的关系时使用标准用语或唯一标识符,实现科研数据之间以及科研数据与相关科研人员、科研机构、科研项目、研究领域、学术出版物及空间位置等各实体间建立丰富的关联关系,从而提升科研数据被发现、被获取和再利用的效果。

#### 4.1 科研数据与科研人员及机构

通过科研数据创建者的背景和影响力来辅助衡量数据的相关性及评价数据质量是一种常用方法,但科研人员的名字及机构名称有多种拼写形式,并

且重名现象也不断加剧<sup>[20]</sup>,即使在组织描述数据时使用了本地标识符,但由于各机构的命名规则不同,或者人员和机构存在多种从属关系,都可能会导致不同标识符同时指向同一个科研人员或机构。因此利用通用的唯一标识符来构建科研数据与相关人员及机构的关联就变得至关重要。ANDS 通过与澳大利亚国家图书馆合作,拓展已有的人员与机构标识符系统(NLA Party Identifier),改善自动和人工的身份匹配服务,从科研部门收割人员和机构信息自动添加到注册系统中,从而更好地支持数据发现。此外 ANDS 也成为了开放研究者与贡献者身份认证机构(ORCID)的成员,促进了对科研人员信息的获取和识别。

#### 4.2 科研数据与科研项目

科研项目会产生大量数据,建立二者之间的关联可以提升数据的发现和再利用的效率。由于大型的科研项目可能是跨机构、跨学科、长期的科研活动,需要使用一个通用的唯一的标识符对这些项目产生的数据进行描述。UKDS、SND 等为每个科研项目分配唯一项目编号,实现了对人口普查等系列调研数据和统计数据的集成和关联。ANDS 在这方面也进行了探索,通过与科研资助机构澳大利亚科研委员会(ARC)以及国家健康与医学研究委员会(NH&MRC)的合作,为其所有资助项目提供持续的、唯一的、可引用的标识符 URI 服务,并且可以在数据发现平台中实现科研数据与科研项目的关联。

#### 4.3 科研数据与研究领域

通过提供科研数据与其所属研究领域的关联,可以实现对数据的精确检索,不仅有利于同领域科研人员的数据共享,同时也为跨学科领域的科研数据发现和利用提供了基础。在调研对象中,澳大利亚、英国、荷兰、瑞典四个国家科研数据服务中都提供了科研数据与研究领域的关联服务,虽然根据所拥有数据的不同,对于研究领域的分类标准并不完全一致,但是基本可以满足用户通过研究领域角度对科研数据的浏览和检索需求。ANDS 为了实现对科研数据更精确的检索,研发了 RIF-CS 受控词表,与澳大利亚统计局和新西兰统计局合作开发了“澳大利亚与新西兰科研领域分类表(ANZSRC)”来描述科研活动中的不同研究领域,该分类标准建立了数据与研究领域的紧密关联,提供了以研究领域为切入点的数据检索和浏览。

#### 4.4 科研数据与学术出版物

从目前的实践情况来看,类似于期刊文献的引用,论文中正式引用科研数据或者数据发布时提供与之相关的出版物等方法,可以有效建立数据与论文之间的关联。通过数据与出版物的互操作,一方面读者可以通过文献快速定位以及获取数据,增加了数据的重用性和共享性,另一方面也可以从数据出发找到与之相关的文献,提升科研活动的创新能力。数据引用机构(DataCite)在推进数据引用规范化方面做出了重大贡献,使用数字对象唯一标识符(DOI)系统实现对科研数据的发布和引用,得到很多国际机构的认可,目前 ANDS、SND、DCC 等 30 多个机构成为 DataCite 的成员,可以独立为数据集分配 DOI,对促进科研数据的规范引用以及实现科研数据与出版物的关联具有重要推动作用。NDS 联合国际化多媒体出版集团爱思唯尔(Elsevier)、自然(Nature)以及高校图书馆和数据中心等机构,开展了探索科研数据与出版物关联项目(OLDRADA),积极寻求以更加开放和灵活的方式实现不同数据仓储库中数据与出版物的关联问题。

#### 4.5 科研数据与空间位置

将科研数据作为公共资源,并从国家层面提供科研数据服务的重要目标之一是促进跨学科的知识发现,其中数据的空间位置是一个重要的关联机制。如果数据集的描述包括空间范围数据,并且是以地理图示而非文本格式描述,数据资源的价值将大幅提升。ANDS 与澳大利亚地球科学组织合作开发了在线本地地名服务项目。该项目通过公开接口提供权威的国家地名辞典和其他有效的空间位置信息,可以协助科研人员、机构或数据存储库生成更多的与数据相关的空间信息,有助于发现新的数据关联及数据融合的机会,进一步促进新型的科研创新活动。

### 5 科研数据的发现

科研数据的发现是指通过对全国范围内海量的、分布式的科研数据进行标准化和规范化的管理,从国家层面建立一个经过整合的科研数据发现平台,使得用户可以克服不同机构、不同学科领域的分散性和差异性,高效便捷地查找到感兴趣的科研数据。

在本文的调研对象中,澳大利亚的科研数据发

表 1 国家科研数据发现服务

国家科研数据服务平台	发现系统	数据量	科研数据发现入口							
			主题	人员	机构	时间	空间	标识符	获取程度	其他
ANDS	Research Data Australia	110887 个数据集	✓	✓	✓	✓	✓	✓	✓	许可协议类型
SND	SND's Collection	1602 个研究项目	✓	✓	✓	✓	×	✓	✓	数据类型 分析单元 研究类型
DANS	EASY	30619 个数据集	✓	✓	×	✓	✓	✓	✓	格式语言
UKDS	UK Data Service Discovery	6500 个数据集	✓	✓	×	✓	✓	✓	✓	数据类型 分析单元 获取工具

注:科研数据的数量统计截止时间为 2016 年 1 月 17 日。其中 SND 的数据是以研究项目为基本单位进行管理的,每个项目下有数目不同的数据集,无法精确统计全部数据集的数量,因此使用研究项目的数量代替数据量。

现平台(Research Data Australia)是覆盖学科范围最全面、功能最完善的平台。截止到 2016 年 1 月份,已经囊括了来自于 100 多个澳大利亚的高校、科研机构及政府机构的 110887 个数据集。该平台是在数据集合注册表的基础上,利用 RIF-CS 的元数据框架、唯一标识符(DOI、PURL、ORCID 等),通过丰富的关联策略实现了科研数据与人员、机构、主题领域等各实体间的关联,帮助用户扩大科研数据发现的范围,增加了科学数据被再次利用的可能性。此外,ANDS 将在未来发现平台中逐步提供数据关系网络图、数据地图接口、云标签等功能,使得用户可以找到意料之外的数据集<sup>[21]</sup>。ANDS 还积极寻求与某些特定学科的数据检索发现服务的互操作,通过链接跳转到具有不同学科特点的服务接口,如化学分子式的发现工具,使用户可以更加便捷、深入、全面地获取所需数据。SND、DANS、UKDS 也通过特定的平台提供科研数据的发现服务,能够有效帮助用户更好地查找和获取感兴趣的数据资源。如表 1 所示,除普遍提供主题、人员、机构、时间、空间、标识符及获取程度等发现入口外,有的平台还提供许可协议类型、分析单元、语种等入口。

NDS 和 RDC 的重要目标之一是帮助用户跨越学科界限进行科研数据的获取。虽然目前尚未建设一个统一的发现平台,但是 NDS 在其年度报告中指出单纯的跨学科检索会非常局限,将深入开发针对特定联盟或仓储库的发现工具作为今后的重要工作内容。

## 6 科研数据的再利用

科研数据的再利用也称为“数据重用”,通常是

指数据被创建及共享后,科研人员可能出于新的研究目的或以不同方式对数据进行重新分析和利用的行为。这种方式一方面有助于再次验证研究成果的科学性,另一方面也可以最大程度发挥数据的价值。科研数据的创建、收集及整理往往需要耗费大量人力、物力和时间,如果缺乏成熟完善的数据再利用的意识和环境,会使得这些来之不易的数据的重用率非常低。国家科研数据服务通过对基础设施和数据资源的集成,可以提供相对完善的数据发现和关联服务,同时也在数据的质量评价、获取权限及知识挖掘等数据的再利用方面提供了相应的指导和支持。

### 6.1 数据的质量评价

科研数据的质量对于整个科研活动是至关重要的,甚至决定了科研的成败,然而如何对科研数据的质量进行评价是一个比较复杂的问题。为了帮助科研人员在利用数据时进行准确筛选和评判,ANDS 发布了指导手册,建议科研人员充分了解数据集本身的信息,包括数据收集过程中使用的工具及软件、样本数据的特点、数据是如何从原始数据演变而来的(例如是否经过加权处理、缺失值的处理方法)、数据集中的变量及变量的结构等等。此外,与传统的论文质量评估方法类似,科研数据的质量评价也包括对信度和效度等因素的评价,如来源的可信性、数据收集的时间跨度和响应率是否合理、数据收集方法在所属研究领域的合理性和可接受程度、数据收集行为是否具有一致性等。NDS 的服务目标之一是“帮助科研人员使用数据”,其中也明确指出获取原始数据以及处理和分析数据时使用的软件、或者数据集的注释信息等内容,可以帮助研究人员理解

表 2 SND 的数据获取级别

级别	说明	具体情况
1	无需与数据提供者或 SND 取得联系即可以获取数据	1a: 无需在 SND 上注册即可获取
		1b: 在 SND 上注册后即可获取
		1c: 通过提交申请, 说明姓名、邮箱地址及使用目的等, 即可获取; 申请并不直接传递给数据提供者, 而主要是用于使用情况统计
2	需要一定的限制条件才能获取数据	2a: 在特定条件下(如学生、校外人员等)向数据提供者提交申请, 得到许可后才可以获取;
		2b: 在任何情况下都需要向数据提供者提交申请, 得到许可后才可以获取;
3	SND 只存储元数据信息, 并不包括数据本身	3a: 需要到其他数据仓储库进行获取
		3b: 需要与数据提供者直接联系进行获取

及评价数据集包含的内容, 对正确使用数据具有重要意义。

## 6.2 数据的获取权限

同其他科研产出一样, 科研数据也存在着权利归属问题, 为了保证数据利用的高效性和合理性, 必须明确数据获取时潜在的限制条件。通常情况下, 科研数据的元数据中会包含获取权限, 有时也会提供产生该数据的项目主页链接或数据管理者的具体联系方式等, 方便使用者了解及查询数据获取的具体限制。

ANDS 将科研数据的获取层级分为开放获取(Open Access)、条件获取(Condition Access)、受限获取(Restrict Access)、其他(Others)、不清楚(Unknown)5 种情况。其中条件获取是指数据在满足特定的前提条件下可以实现获取和再利用, 如支付一定的费用或与数据监护者直接联系等。受限获取是指数据在某些限定情况下才可以获取, 如只有特定团队的成员或者特定的地理位置等。

SND 提出了更加细化的数据获取级别, 如表 2 所示。

## 6.3 数据的分析挖掘

科研数据中蕴含巨大的价值, 随着数据分析挖掘技术的不断发展, 这些潜在价值可以被充分发掘, 然而并不是所有的科研人员都具备相应的技术和能力, 这就需要国家科研数据服务提供更高质量和更深层次的增值服务内容。比如推荐或集成科研数据分析、挖掘和可视化呈现的工具软件。NDS 开展的数据展示与分析项目(yt-Hub)对相关服务内容进行了积极探索, 该项目可以提供强大的可视化分析工具包, 为科研团体提供了个性化、多样化的科研数

据展示方案, 还利用云计算技术提供了先进的访问接口, 特别是可以利用 Python 实现对远程数据的自定义分析。UKDS 在 ESRC 的资助下开展了二次数据分析项目(SDAI), 旨在通过对科研数据深层次的知识挖掘来开展一些创新性研究, 从而推进科研数据在更广泛意义上的共享。SND 提供了对科研数据的在线分析功能<sup>[22]</sup>, 目前可以实现对国际社会调查项目(ISSP)、欧洲社会调查项目(ESS-survey)等数据的在线分析。

## 7 合作机构

为了实现跨机构、跨学科、跨领域的科研数据共享和再利用, 国家科研数据服务需要吸引政府机构、数据基础设施、公共文化机构、科研机构、出版机构等相关利益群体参与(如图 2 所示), 充分发挥各自在政策、技术、资源、服务方面的优势, 实现数据、系统、标准、管理的多维度协同, 从而推动科研创新活动的发展。

政府部门是数据资源的最大持有者, 发达国家已经认识到政府数据本身具有的战略价值, 并从政策、立法、技术、产业等各层面对于政府开放数据共享进行引导<sup>[23]</sup>。国家科研数据服务也在争取通过与政府部门的合作, 使得科研人员可以最大程度的获取相关数据, 实现政府开放数据的价值最大化。ANDS 开展了公共机构数据项目(Public Sector Data), 旨在更好地提升公共部门数据的可发现性, 促进科研社区与联邦政府、地方政府的合作。目前已与澳大利亚工业部、统计局、气象局等多个国家政府部门建立合作关系。此外瑞典国家统计局、荷兰国家统计局、加拿大工业部及国家统计局等都纷纷与

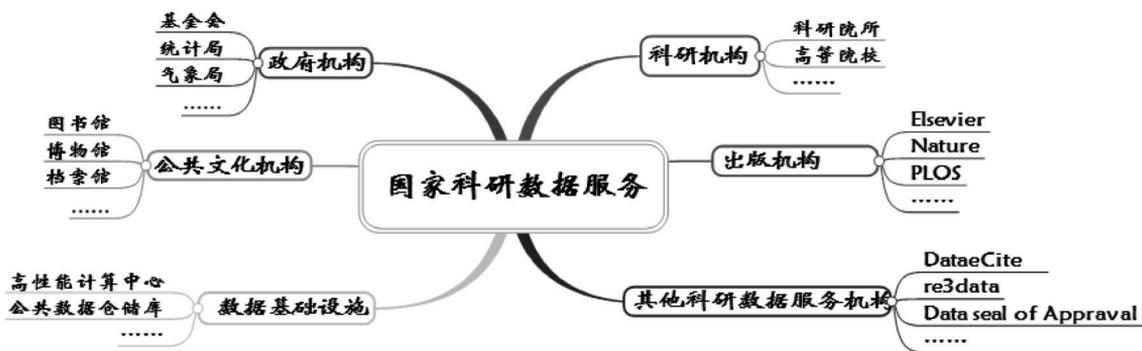


图2 国家科研数据服务合作机构

各国国家科研数据服务合作。基金会作为政府资助科研活动的重要组织,有权对项目所产生数据的提交和共享制定约束政策<sup>[24]</sup>。通过与基金会建立合作,可以更好地整合已有的服务资源,并从国家层面对数据的管理与共享政策产生影响。目前澳大利亚、加拿大、荷兰、英国等都与本国的基金会建立了密切的联系。

作为国家重要的数据基础设施,数据仓储库已经保存了大量政府资助的重大课题、专题研究的数据成果。国家计算中心所具备的高性能计算、大数据处理等能力是实现国家科研数据服务的必要保证。其中澳大利亚数据存储中心(Australian Data Archive)、英国数据存储中心(UK Data Archive)、荷兰数据存储中心(CentERdata)、瑞典国家计算中心(SNIC)、美国国家计算中心(NCSA)、加拿大计算中心(Compute Canada)、以及荷兰计算中心(Netherlands eScience Center)等都是目前各国国家科研数据服务的重要参与者。

国家科研数据服务也积极寻求与图书馆、档案馆、博物馆等公共文化机构的合作,充分利用其海量的数字资源及资料组织、保存与管理方面的经验和优势,为科研数据提供支持和服务。瑞典国家图书馆、档案馆,澳大利亚国家图书馆、档案馆、博物馆等已经成为国家科研数据服务的合作机构。

科研机构是科研数据的主要创建者和使用者,也是国家科研数据服务的主要服务对象。国家科研数据服务希望通过与更多科研院所、科研项目及学科联盟建立密切的合作关系,更好地提升服务覆盖度及影响力。ANDS 与全国四十多所高校保持合作关系,目前 NDS 的合作对象中包括了康奈尔大学、普渡大学等七所高校。

作为科研数据传播交流的媒介之一,诸多国外学术出版社及期刊纷纷对科研数据的存储以及传播提出明确的政策<sup>[9]</sup>。数据出版作为一种新型的出版模式,是将科研数据整合进学术记录,使得科研人员的劳动成果得到学术界和社会认可的重要环节。DNAS 与出版社合作,发布了人文社科领域的科研数据期刊 *E-data & Research*,为研究人员和学生提供关于科研数据存储与再利用的相关信息。目前 Elsevier、IEEE、Nature、PLOS 等出版机构已经确定与 NDS 建立合作关系,在科研数据与科研论文的关联出版、数据论文的出版等方面进行了积极的探索。

随着科研数据逐渐成为人类共同的知识资产,解决日趋复杂的问题不仅需要充分整合国内各学科和领域的资源,有时甚至需要跨国界的协同工作。各国也非常重视与其他国家科研数据服务的联系与合作,ANDS、DANS、RDC 等纷纷建立了合作关系,同时还与数据引用机构(DataCite)、科研数据知识库(re3data)等国际组织加强合作。科研数据联盟(RDA)的成立也为全球的数据共享奠定了基础。

## 8 结论

通过调研的 6 个国家科研数据服务可以看出,这些国家在科研数据管理与服务方面已经取得了一定的进展,特别是在国家层面整合数据资源,提供更高质、更精准、更智能的服务方面进行了积极的探索和实践,对于该领域的后续发展提供了借鉴。

相比之下,当前我国的科研数据服务主要集中在基础科学领域,并且主要提供以学科为导向的数据资源集成服务。最有影响力的项目是国家科学技术部于 2002 年实施的科学数据共享工程,作为国家

科技基础条件平台建设的组成部分,该项目在整合分散的科学数据资源等方面进行了探索,目前已经对环境、农业、人口与健康等共 24 个领域开展了科学数据共享工作<sup>[25]</sup>。在这十多年中,科学数据共享工程的推进在构建国家科研数据管理与共享服务体系、增强科研数据资源积累、促进科研数据增值等方面具有重要贡献。武汉大学、复旦大学及北京大学等也开展了高校层面的科研数据管理与服务机制的实践与探索。但总体来说,我国已有的科研数据服务普遍存在政策法规不完善、缺乏规范的元数据标准、数据可获取性较低、提供的服务形式较少、各个平台的科研效能差距较大等诸多问题<sup>[26]</sup>,缺乏多个平台之间的集成和整合,不利于研究人员跨机构、跨学科、跨领域的科研数据发现与共享。

鉴于国外已有的实践经验,对我国开展国家科研数据服务提出如下建议:

(1)政府加大引导和支持力度,保障科研数据的共享利用。

政策是实施数据管理和服务的重要推动力,国外经验表明,政府为支持和促进数据的广泛共享,逐步将科研数据管理提升到国家战略高度,制定了相应的计划、法律法规和政策。我国政府虽然也先后制定了《科学数据共享条例》、《国家科技计划项目科学数据汇交办法》等一系列数据共享的政策条例,但缺少相应的法律效力,限制了科研数据的有效共享利用,因而需要我国政府加大对科研数据服务的引导和支持,制定并完善相关法律法规,使科研数据服务做到有法可依。

(2)出台相应的数据管理政策,增强科研人员的管理与共享意识。

国外越来越多的科研资助机构要求受资助者在申请科研资助时提交相应的数据管理计划,并阐述将如何对科研过程中产生的科研数据进行管理。目前我国的主要科研资助机构如国家自然科学基金委员会、全国哲学社会科学规划办公室等都没有强制要求申请者提交科研数据管理计划,这使得我国科研人员缺乏数据保存和管理的意识。我国应借鉴国外的成功经验,出台相应的数据管理政策,将科研人员对数据的管理与共享作为强制性的要求纳入考核范围,这将更好地增强科研人员的科研数据管理与共享意识,保证数据保存的规范化与长期化,保证科研投入效益最大化。

(3)积极推进数据仓储库的建设,为实现全国范围数据资源共享奠定基础。

经调研发现,国外许多高校为响应科研资助机构关于数据管理政策的规定,纷纷建设了本机构的数据仓储库,并制定了相应的高校数据管理政策,在帮助科研人员申请科研资助的同时,也实现了对本机构海量科研数据的有效保存与共享。中国科学院在国内率先提出了建设科学数据库的设想,并建设了多个专业数据库,此外武汉大学、复旦大学、北京大学也建设了具有本校特色的科研数据共享平台,但是相较于国外的发展情况,国内机构数据仓储库的数量和质量还远远不够,各科研机构应利用各自特色数据资源建设机构数据仓储库,开发科研数据服务的相关软件工具,积累科研数据服务实践经验,同时注重遵循统一的标准,充分利用国际化的开源软件,为实现全国范围数据资源共享服务奠定基础。

(4)深化科研数据服务层次,提升科研数据服务质量。

科研数据来源于科学研究,同时也服务于科学研究,是未来科技创新的重要资源。当前已有的国家科研数据服务都非常重视对跨机构、跨学科、跨领域的科研数据的发现、关联及知识挖掘,并取得了显著的进展。我国的科研数据服务也应借鉴国外的成功经验,实现对全国范围内不同科研机构、学科领域的分布式数据资源的整合与统一访问,同时可以集成科研数据分析、挖掘和可视化呈现的工具软件,面向特定学科领域与应用提供特色增值服务,逐步深化科研数据的服务层次,增强科研数据的被发现、被获取和再利用的效果,提升科研数据服务质量。

(5)积极寻求国内外相关利益群体的合作,增强科研数据服务能力。

随着科研数据逐渐成为科研协同创新的重要资源,解决日趋复杂的问题一方面需要吸引政府机构、数据基础设施、公共文化机构、科研机构、出版机构等相关利益群体的共同参与,促进国内各学科和领域的资源的充分整合,另一方面要与其他国家科研数据服务建立紧密的联系,学习和借鉴其成功经验,避免走错路、走弯路。此外,还要积极参与到科研数据联盟(RDA)、科研数据引用(DataCite)、科研数据知识库(re3data)等国际组织中,提升自身服务优势,开发新型的数据服务项目,增强科研数据服务能力。

2015年的科技数据资源整合与共享工程总结

会议指出,要推动国家科学大数据中心建设,突出“应用牵引,统筹兼顾”的思路,加强科研数据服务体系的顶层设计,推动科研数据共享。我国需要在借鉴学习国际成功经验基础上,宏观规划国家科研数据服务支撑项目,有效整合各机构和各学科领域的科研数据资源和基础设施,深化科研数据服务层次,最终构建具有中国特色的国家科研数据服务体系。

### 参考文献

- 1 Katherine G. Building support for research data management: biographies of eight research universities[EB/OL]. [2015-10-10]. <http://www.ijdc.net/index.php/ijdc/article/view/327>.
- 2 Y Shen, V E Varvel. Developing data management services at the Johns Hopkins University[J]. *Journal of Academic Librarianship*, 2013, 39(6): 552-557.
- 3 杨云秀,顾立平等. 国外科研教育机构数据政策的调研与分析——以英国10所高校为例[J]. *图书情报工作*, 2015(5): 53-59.
- 4 A M Cox, S Pinfield. Research data management and libraries: current activities and future priorities[J]. *Journal of Academic Librarianship and Information Science*, 2014, 46(4): 299-316.
- 5 杨鹤林. 数据监护:美国高校图书馆的新探索[J]. *大学图书馆学报*, 2011(2): 18-21.
- 6 彭建波. 北美人文社会科学数据管理的实践及其启示[J]. *大学图书馆学报*, 2013(6): 33-37.
- 7 Jennifer L. Make data sharing easy: PLOS launches its data repository Integration partner program[EB/OL]. [2015-10-10]. <http://blogs.plos.org/tech/make-data-sharing-easy-plos-launches-its-data-repository-integration-partner-program>.
- 8 吴蓉,顾立平,刘晶晶. 国外学术期刊数据政策的调研与分析[J]. *图书情报工作*, 2015(7): 99-105.
- 9 Carr D, Littler K. Sharing research data to improve public health: a funder perspective[EB/OL]. [2015-10-10]. <http://www.ncbi.nlm.nih.gov/pubmed/26297752>.
- 10 张瑶,顾立平,杨云秀等. 国外科研资助机构数据政策的调研与分析——以英美研究理事会为例[J]. *图书情报工作*, 2015(6): 53-60.
- 11 Lyon L. Dealing with data: roles, rights, responsibilities and relationships[EB/OL]. [2015-10-10]. [http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing\\_with\\_data\\_report-final.pdf](http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf).

- 12 樊俊豪. 图书馆在科学数据管理中的角色定位研究[J]. *图书情报工作*, 2014(6): 37-41.
- 13 Background to the 2011 Canadian Research Data Summit [EB/OL]. [2015-10-10]. <http://www.rdc-drc.ca/wp-content/uploads/Background-to-the-2011-Canadian-Research-Data-Summit1.pdf>.
- 14 Adding value by taking a national and institutional approach to research data: the ANDS experience[EB/OL]. [2015-12-20]. <http://www.ijdc.net/index.php/ijdc/article/view/8.2.89>.
- 15 National science foundation [EB/OL]. [2015-12-20]. <http://www.nsf.gov/index.jsp>.
- 16 National collaborative research infrastructure strategy (NCRIS) [EB/OL]. [2015-12-20]. <https://education.gov.au/national-collaborative-research-infrastructure-strategy-ncris>.
- 17 Data conservancy [EB/OL]. [2016-04-07]. <https://dataconservancy.org>.
- 18 Background to the 2011 Canadian Research Data Summit [EB/OL]. [2015-12-20]. <http://www.rdc-drc.ca>.
- 19 Berman F, R Wilkinson, J Wood. Building global infrastructure for data sharing and exchange through the research data alliance [J]. *D-Lib Magazine*. 2014, 20(1-2).
- 20 黄国彬,郑琳. 科研人员唯一标识符的组成与应用研究[J]. *图书情报工作*, 2015(4): 25-31.
- 21 A Burton, AETreloar. Designing for discovery and re-use; the ‘ANDS Data Sharing Verbs’ approach to service decomposition[J]. *International Journal of Digital Curation*, 2009, 4(3): 44-56.
- 22 SND online service [EB/OL]. [2015-12-20]. <http://analysis.snd.gu.se/webview/?language=en>.
- 23 张涵,王忠. 国外政府开放数据的比较研究[J]. *情报杂志*, 2015(8): 142-146.
- 24 樊俊豪. 图书馆在科学数据管理中的角色定位研究[J]. *图书情报工作*, 2014(6): 37-41.
- 25 中国科技资源共享网[EB/OL]. [2015-12-20]. <http://www.esi.cn/metadata/science.jsp>.
- 26 司莉,李月婷等. 我国科学数据共享平台绩效评估实证研究[J]. *图书馆理论与实践*, 2014(9): 30-35.

作者单位:北京师范大学政府管理学院,北京,100875

美国德雷塞尔大学计算机信息学院,费城,19019

收稿日期:2016年5月20日

## The Progress and Enlightenment of National Research Data Services

Liu Xiaojuan Yu Jia Lin Xia

**Abstract:** This paper introduces the background and present situation of National Research Data Services of several countries including the United States, United Kingdom, Canada, Australia, Netherlands and Sweden. It summarizes systematically the practical experience of these services in terms of infrastructure construction, data connection, data discovery, data reuse and cooperative agency. Research shows that these countries have made some progress in national data service. In contrast, China's research data services are mainly concentrated in the field of foundation subjects, and mainly provide a subject oriented data resources integration service. We need to increase the government's guidance and support, plan national research data service projects, effectively integrate scientific data resources and infrastructure of various institutions and disciplines, deepen the data service level, and finally build up a national data service system with Chinese characteristics.

**Keywords:** National Research Data Service; Data Connection; Data Discovery; Data Reuse