

面向科学数据管理的机构知识库 建设与启示*

唐玫 赵婉忻 李晶 师俏梅
(西北工业大学图书馆, 西安 710072)

摘要: 开放科学背景下科研活动愈发密集, 科学数据管理愈发重要, 传统机构知识库立足于数据仓储的基本功能, 正在逐渐向科学数据管理和知识增值服务拓展。本文在调研面向科学数据管理的国内外机构知识库建设现状的基础上, 论述其建设的关键技术, 并以西北工业大学机构知识库服务实践为例, 提出面向科学数据管理的机构知识库建设与发展的建议: 优化管理政策、制定科学数据管理规范、嵌入科学数据生命周期、组建专业科学数据管理团队以及加强平台间互联合作。

关键词: 开放科学; 科学数据管理; 机构知识库

中图分类号: G251.5 **DOI:** 10.3772/j.issn.1673-2286.2022.10.012

引文格式: 唐玫, 赵婉忻, 李晶, 等. 面向科学数据管理的机构知识库建设与启示[J]. 数字图书馆论坛, 2022(10): 65-72.

当前全球科研范式发生着深刻的变革, 科研活动更加密集, 其产生的科研信息及其存储、传播和交流都发生着巨大变化。科学数据 (Research Data) 也称研究数据, 是人类社会科技活动中所产生的基本数据和按照不同需求系统加工的数据产品及相关信息^[1], 涵盖了科研活动过程中产生的可被数字化存储的各类信息。科学数据管理 (Research Data Management, RDM) 服务指在科学数据组织、长期保存与共享等管理活动中开展的各类服务^[2]。传统机构知识库仅作为机构数据的仓储中心, 服务模式和服务内容在日渐复杂的科研活动中发展受限。鉴于此, 开放获取知识库联盟 (Confederation of Open Access Repositories, COAR) 于2017年启动并持续开展“下一代机构知识库” (Next Generation Repositories, NGRs) 的建设, 为机构知识库提供改进功能的方式、协议和技术支持, 其建设目标是在传统机构知识库构建框架上部署知识增值服务并将其提升为能面向研究、开放、有助于创新、同时便于学术群体集体管理的平台^[3]。2021年, 我

国教育部发布《高等学校数字校园建设规范 (试行)》, 明确鼓励高校将机构知识库建设与科研资源管理相结合, 为机构知识库的转型提供了理念和思路。正是由于机构知识库建设存在脱离科研流程、成为存储库、最终走向“信息孤岛”的痛点, 因而提倡机构知识库从建设之初就与科研资源管理相结合, 对科学数据进行采集、汇交、保存、管理、共享, 服务于科研活动全流程, 有助于形成流动、完备的科研生态体系。

无论从科学数据管理还是机构知识库的需求出发, 传统机构知识库都有必要进行技术的革新和服务内容的转型。在开放科学环境下, 面向科学数据管理提供服务是机构知识库发展的必需, 也是未来科研活动的大势所趋。

1 相关文献研究

在科学数据管理领域, 美国、英国和澳大利亚等国家的相关工作较为成熟, 涉及科学数据管理的概念界

* 本研究得到西北工业大学发展战略研究基金“‘双一流’建设背景下我校战略规划现状与提升研究” (编号: 2022FZY42) 资助。

定、政策制定、平台建设及服务实践等,国内学者分别就这些方面开展了较为丰富的研究。Neil Beagrie界定了科学数据的概念^[4];张玉娥等^[5]、周晓燕等^[6]、黄国彬等^[7]分别对欧盟、澳大利亚、英国的科学数据管理政策进行调查和研究;魏悦等^[8]调研了美英澳三国的15所高校科学数据管理政策内容,构建了基于生命周期理论的科学数据管理与共享政策内容框架;袁晓明等^[9]调研了芝加哥大学基于微服务架构的Globus平台应用于科学数据管理服务的实例;汪满容等^[10]从计划、策略、管理与保存等方面调研了4所世界知名的新加坡高校图书馆针对科学数据管理的服务模式;王继娜^[11]从栏目设置、服务方式、服务内容、岗位设置等方面对英国、德国、新西兰等国家的20所高校图书馆数据管理服务实践进行归纳总结;张贵香等^[12]认为高校科学数据知识库及其联盟的建设是有必要的。

在机构知识库方面,学界认为以2007年为界限,机构知识库由快速发展期转变为稳定增长期,研究热点由机构知识库的建设转变为提升使用度、调整建设模式、扩展服务内容和深度开发功能的可持续发展^[13]。Wilson等^[14]早在2011年牛津大学科学数据管理基础设施系列项目中提出要以研究人员为中心、坚持科学数据管理服务的机构知识库建设目标;Kim等^[15]认为机构知识库可通过网页门户服务共享该领域学者产生的科学数据;陈和等^[16]对中文期刊论文中的机构知识库主题进行数据挖掘,认为2017年后机构知识库的研究热点为科学数据、科研服务、辅助科研决策和数据增值服务等,这与吴建中^[17]对科学数据管理工作中要解决的问题一致,他认为科学数据管理需要融入机构知识库;崔海媛等^[18]论述了“下一代机构知识库”的关键技术、主要功能和发展趋势;都平等^[19]对“下一代机构知识库”的资源类型和特点进行了具体研究;王思敏等^[20]从总体架构、服务流程和数据服务等方面构建了一套服务于科研管理的精准型机构知识库。

当前,分别从图书馆视角进行科学数据管理的理论研究及机构知识库服务扩展的研究已然较为成熟,但将科学数据管理与机构知识库进行融合的相关理论及实践研究仍处于初步探索阶段,尚未普及。因此,本文在调研面向科学数据管理的国内外机构知识库建设现状的基础上,论述其建设的关键技术,并以西北工业大学机构知识库服务实践为例,为推动面向科学数据管理的机构知识库建设与发展提出建议。

2 面向科学数据管理的机构知识库建设现状

2.1 国外现状

欧盟通过“地平线2020计划”(Horizon 2020)明确提出受其资助的项目产生的科学数据需要实现开放获取,欧盟机构知识库建设已普遍向科研服务、科研评价、决策支持等方面转型,同时开展与传统的科研信息管理系统(Current Research Information Systems, CRIS)的衔接研究,并开发和启用了相应的互操作平台,满足科研机构开展信息管理的需求。英国支持的Publisher & Institutional Repository Usage Statistics项目,研究利用COUNTER指标统计机构知识库内容的使用情况,支持科研管理者对项目成果进行统计评价^[21];爱尔兰都柏林圣三一大学的机构知识库(TARA Institutional Repository)独立作为学校科学数据开放获取的平台,支持面向搜索引擎的完全开放和检索,鼓励学者自主缴存其在科学活动中产生的科学数据^[22];荷兰的国家研究与合作信息系统(NARCIS门户)^[23]实现了机构知识库与科研系统信息管理的集成;圣安德鲁斯大学(Universitiy of St Andrews)机构知识库与科研信息管理系统通过API接口实现互联和互操作^[18];伦敦国王学院(King's College London)在科研信息管理系统中嵌入了具有开放获取功能的机构知识库模块^[18];德国马克斯·普朗克科学促进会(Max Planck Gesellschaft, MPG) eScidoc知识库^[21]对科研活动各阶段产出的对象进行语义标记和管理,实现机构知识库与科研过程的结合和互操作,支持科研工作流,进行科研评价;拥有662家机构成员的日本机构知识库联盟JPCOAR于2020年10月发表了针对研究人员的科研发展管理(Research & Development Management)教材,加入联盟的机构知识库嵌入了科学数据公开、管理、共享和评估的功能^[24]。

2.2 国内现状

香港大学学术库(HKU Scholars Hub)是香港大学学术资源总库,基于DSpace系统框架,在学术库上集成了科研信息管理模块,与科研成果系统集成并实现互操作,同时从教务处、科研事务部、研究生院、公共事务处等子系统处共享学者科研项目、基金、专利、荣誉

及奖励等,开展科学数据服务的拓展^[25]。北京大学机构知识库于2015年升级至3.0版本,2018年基于DSpace框架使其嵌入北京大学科研信息管理系统,同时引入代替计量学理念,增加机构知识库的学术评价功能;此外,2015年上线的基于Dataverse开源系统开发的北京大学开放研究数据平台与机构知识库间也实现互操作,使得北京大学科学数据通过平台存储于机构知识库中^[18]。西北工业大学知识管理系统与上海科技大学知识管理系统项目基于中国科学院的Cspace架构,收集并存储学校科学数据成果,通过API接口与校内其他系统进行互访问,向各业务部门提供学者的各种成果数据,不再允许各业务系统或考核流程单独采集成果数据。除此之外,目前大部分机构知识库还定位于提供知识仓储的功能,用户黏度欠佳。

科研人员的对接、科研平台的整合、科学数据管理和共享方式都对当前高校及科研机构的科学数据管理提出了新的需求和考验。当前,一部分机构知识库已经正视科学数据管理的服务需求,一定程度上增加了数据整合的拓展功能,但是绝大部分仍在扮演机构数据仓储的角色,在顶层设计阶段未考虑科学数据管理的服务需求,且存在不对外网开放、平台架构及功能设计不完善、数据长时间未更新、未能整合资源等弊端,在科研活动中能为科学数据管理提供的服务非常有限。

3 面向科学数据管理的机构知识库建设的关键技术

近年来,考虑到科学数据管理的必要性,各国出台了一系列规范化的政策文件用来指导科学数据管理工作。2018年国务院办公厅印发的《科学数据管理办法》(以下简称《办法》)从科学数据采集、汇交、保存、共享及利用等方面提供了国家层面针对科学数据管理的规范化指导^[26]。笔者参考《办法》并分析开放获取知识库联盟下一代机构知识库工作组研究报告《COAR行为及技术建议》^[3],梳理出机构知识库为面向科学数据管理不同阶段的要求,在传统机构知识库缴存功能基础上须采用的关键技术。

(1) 科学数据采集与汇交。《办法》规定科学数据生产者要按照相关标准规范组织开展科学数据采集生产和加工整理,形成便于使用的数据库或数据集,同时应建立科学数据汇交制度,为政府资助的各项科技计划(专项、基金等)项目各阶段形成的科学数据、

学术论文等汇交到对应机构。机构知识库须采用的关键技术包括:公开标识符(Exposing Identifiers)、强调为元数据赋予唯一标识符,从源头上部署规范化采集数据的技术;添加明确的许可标识、声明许可协议(Declaring Licenses at a Resource Level),同时对许可协议、权益政策进行集中管理,便于不同类型数据规范化采用;提供用户识别(Identification of Users)和用户认证(Authentication of Users)功能,支持ORCID与系统互联等用户交互功能,自动连接其他平台的学术成果;提供资源长期保存(Preserving Resources),通过标准、协议和互操作,建立数据的长期保存机制。

(2) 科学数据管理与共享。《办法》要求须为政府资助形成的科学数据进行开放、分级分类并组织汇编科学数据目录,接入数据共享交换平台。机构知识库须采用的关键技术包括:针对科学数据可依据研究专题、成果类型、作者、完成时间、学科分类等多维度组织并管理数据,为数据提供目录制定和分类导航;采用开源的技术框架,实现与其他科研系统的互访问和互操作;同时具备资源交互功能(Interacting with Resources),支持用户基于第三方社交媒体对科学数据进行注释、分享和评论,有助于提升科学数据的影响力和关注度。

(3) 科学数据增值与评价。《办法》指出须对科学数据进行分析挖掘,形成有价值的科学数据产品,开展增值服务;同时建立完善有效的激励机制,组织机构形成科学数据的评价办法,对科学数据工作进行评价考核。机构知识库须采用的关键技术包括:公开标准化计量指标(Exposing Standardized Usage Metrics),在传统机构知识库的基础上,采用通用标准协议、方法和互操作手段开发和部署数据评价的增值功能,嵌入Altmetrics等计量指标并致力于发展新的数据评价体系。

4 面向科学数据管理的机构知识库服务实践——以西北工业大学知识管理系统为例

4.1 服务需求分析

当前,科研活动的密集与复杂使得科学数据日益趋于海量、复杂化、分散化、时效化以及难以精确化^[27],科学数据管理被赋予越来越重要的意义。随着

近年来图书馆服务向科学数据管理扩展的趋势,高校职能部门及科研人员面向科学数据管理的服务需求包括:①学校办公室等职能部门对全校科学数据的总体态势有查找、统计、总览、监测、分析、可视化等需求;②科研管理部门对科研人员科研活动全流程中涉及的科学数据产生、保存、分类、管理、归档等管理需求;③科研人员对科学数据及相关科研成果缴存和维护的需求。

面对以上服务需求,图书馆通常承担部分科学数据统计、保存和分析等工作。然而,由于缺乏统一规范和标准化科学数据管理平台,馆员开展科学数据的统计、存储和分析的工作模式,使得科学数据归档和再利用往往遇到瓶颈和困难,同时不利于科学数据的长期保存;针对不同职能部门在不同时段对不同科学数据统计和分析需求,容易造成负责馆员重复交叉劳动,使得工作效率降低。因此,有必要采用现有系统平台协助馆员开展科学数据管理工作。

4.2 面向科学数据管理的机构知识库服务实践

西北工业大学图书馆依托中国科学院机构知识库

的Cspace架构,开发了西北工业大学知识管理系统,一期工程于2021年9月正式上线,全面代替原西北工业大学机构知识库。在保留传统功能的基础上,系统结合科学数据管理服务需求,对科学数据的元数据描述、数据采集、储存、分类、共享、交换、分析挖掘、安全与公开等方面开发了相应功能,并通过了测试,为科学数据管理工作提供平台和功能支持。

4.2.1 面向科学数据管理的知识管理系统服务架构

笔者依据以往开展的服务内容和未来服务需求,采用资源层、应用层、服务层和用户层的结构来描述知识管理系统面向科学数据管理工作的服务架构(见图1)。其中,资源层用于描述知识管理系统所集成的现有资源,应用层用于描述知识管理系统中与校内其他子系统间资源互联的模式,服务层用于描述知识管理系统面向科学数据管理时针对科学数据的采集、汇交、保存、共享、交换、分析与评价各阶段中可提供的各项数据服务,用户层描述现阶段知识管理系统的服务需求和系统使用对象。

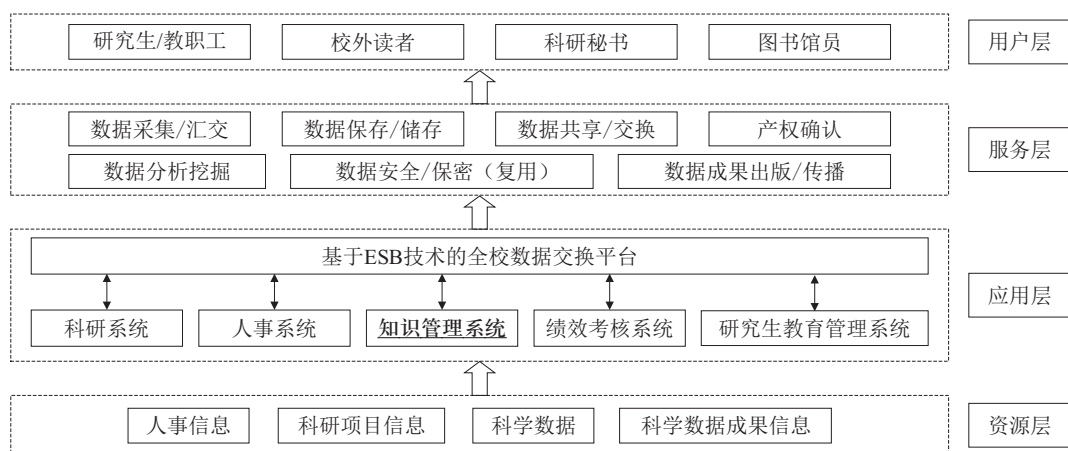


图1 面向科学数据管理的知识管理系统服务架构

4.2.2 面向科学数据管理的系统服务功能

基于系统服务架构,从科学数据管理的各个阶段阐述西北工业大学知识管理系统面向科学数据管理的功能设计和服务模式。

(1) 科学数据采集/汇交。知识管理系统面向全校学者,并嵌入学校统一认证登录方式,为每位学者提供

个人工作间,支持学者自主上传个人科学数据及相关成果。学者可自主申请缴存科学数据(基于SWORD协议方式),经自主查重、管理员审核无误后方可缴存入库。除学者自主缴存外,负责系统数据维护的管理员也可实现外部科学数据的批量导入,为此,知识管理系统为学院科研秘书制定和开启“学院数据管理员”权限,使得学院数据管理员与图书馆馆员共同完成院系科学数据

的管理。此外,该系统也可通过互操作功能,集成其他数据平台的科学数据,完成科学数据汇交。

(2) 科学数据保存/存储。为实现科学数据保存,知识管理系统依照元数据规范设计了文献类型、所属单位、题名、作者、作者单位、发布/更新日期、简介、版本、资助项目、URL等23个字段,其中“数据集类型”将科学数据的类型划分为5种,即实验数据、观测数据、仿真数据、汇编数据和派生数据,以此形成描述科学数据的元数据模板。该系统为一组科学数据分配唯一的统一资源标识符,支持数据的长期储存和检索。笔者以文献[28]撰写过程中的两份原始数据集为例,依据原始数据类型提交并在个人工作间中分别保存了一套仿真数据和一套汇编数据,并分别由题名、作者、作者单位、发布日期、数据集类型、简介、关键词、学科领域、创建日期、更新日期与条目标识符等重要字段进行具体描述。

(3) 科学数据共享/交换。知识管理系统提供标准的REST API数据开放接口,通过基于ESB(Enterprise Service Bus)技术的数据交换平台,与学校其他重要子系统数据进行数据共享和交换。知识管理系统中学者的人事信息及院系归属获取自学校人事系统,并保证随时同步更新,免去学者因人事变动而发生的更新滞后和错误,确保科学数据所属学者信息的准确性;此外,与研究生教育管理系统、绩效考核系统等实现数据的互联,为数据的交换及互操作提供便利。

(4) 科学数据产权确认。对于通过管理员批量汇交或者通过系统间共享方式采集得到的新入库科学数据,通过遍历搜索学者姓名以完成自动产权匹配。知识管理系统集成每位学者约40种别名形式,同时支持学者自主手动添加别名;同时支持学者对数据的产权确认以及对错误匹配的在线申诉及修改。多种途径建立同名消歧机制,以此有效提高科学数据产权确认率,确保科学数据的后续使用。

(5) 科学数据分析挖掘。支持成果按照多种类型(文献类型、发表日期、院系、重要成果、收录类别、学校署名类型、语种等)进行检索及统计,并支持成果共计60余项字段导出,供管理员二次数据加工、科研绩效统计分析、自动生成ppt、word规范形式的可视化报告。同时支持学者在前台根据科研需求进行自主成果分析。

(6) 科学数据安全/保密(复用)。知识管理系统支持多种共享协议模板,学者可自行选择科学数据开放的范围,笔者在本例中采用“知识共享署名-非商业

性使用-相同方式共享许可”(CC BY-NC-SA)协议开放共享数据集,访问权限范围内的访问者可以下载并浏览数据集,便于其他学者后续对数据进行复用或进行科研交流。

(7) 科学数据成果出版与传播。学者在科学数据复用和成果出版过程中,形成学术论文、专利、著作、学术报告等众多科学数据成果资源,这些成果资源的保存、管理和共享是机构知识库的基础功能。当前知识管理系统中已收录各类科学数据成果资源约23万条,其中14万条资源已至少有一名作者完成产权确认。该系统作为校内论文成果统计的唯一来源,通过接口为人事系统、绩效考核系统、研究生教育管理系统等开放论文数据共享。

4.2.3 服务成效与不足

西北工业大学机构知识库(知识管理系统)为科学数据的采集、汇交、保存、共享、交换、分析、成果的出版与传播提供平台,并且支持为研究生教育管理系统提供研究生科学数据的论文成果数据,支持研究生学术生涯各阶段学术成果的统计与管理。截至2022年9月,系统总访问量达118万次,其中,2021年11月和12月,2022年3月、4月和6月访问量均超过10万次,约是其他月份访问量的1.5倍,与学校的科研活动(博硕士生学位申请、教师绩效考核)活跃度成正比,表明该系统的用户黏性很可观、在科研活动中逐渐成为必不可少的平台与工具。

知识管理系统在建设过程中考虑了科学数据管理服务需求,当前从架构和功能角度已然实现对科学数据管理过程的服务支持。然而,由于科学数据管理工作在理论和实践中均处于初级阶段,科学数据管理政策尚未健全、科研人员数据保护意识较强、平台宣传不到位等原因使得本校科学数据体量有限,绝大部分科学数据仍分散存储于科研团队和科研人员手中,尚未得到统一的管理;同时,数据保密、数据权益、数据版权归属以及数据评价等功能有待完备。随着科学数据管理工作的进展,机构知识库面向科学数据管理服务还有细化和健全的空间。

5 面向科学数据管理的机构知识库建设的思考与建议

对高校和科研机构来说,未来如何建设机构知识

库,使其一方面有效地服务于科学数据管理,另一方面提升机构知识库在科研活动中的重要性,向“下一代机构知识库”转型,在顶层政策、服务模式、团队建设以及共享互联等方面都面临重重挑战,以下4点建议是在西北工业大学图书馆知识管理系统建设经验基础上进行的思考和总结,是高校机构知识库建设今后需要努力的方向。

5.1 优化管理政策,制定科学数据管理规范

据Open DOAR统计,截至2022年7月19日,全球机构知识库收录的资源类型有12种,其中科学数据及相关灰色资源有6种,占比37.54%,以波士顿大学为代表的国外高校机构知识库,包含大量会议材料、数据库、软件、乐谱、法律备忘录等形式多样的资源,资源类型多达28种^[29-30];而国内这一比例仅为不到10%,原因在于科研人员存储意向不明朗^[29]。

当前我国以《办法》为代表的政策性文件已从宏观上强调科学数据管理工作的必要性,并用以指导科学数据管理工作的开展,但高校及科研机构内部在宏观政策文件的指导下还需形成和健全本单位内针对科学数据管理的制度、规范和细则。首先,科研机构应明确科研人员对科学数据管理的需求,然后针对当前科研人员在机构知识库中缴存意愿不积极的现状,联合如人事部门、科研管理部门、研究生管理部门等其他单位,将机构知识库的科学数据收录与维护的情况纳入考评流程,建立兼具强制性和激励性的缴存政策,提高科研人员对科学数据管理的认同及缴存的积极性。其次,应优化管理政策:在制定机构科学数据管理制度时,需考虑到数据安全、数据保密、数据归属、数据伦理等问题;在制定科学数据管理汇交制度时,需对科学数据质量、格式和内容进行鉴别;在制定科学数据的分级分类制度时,需考虑如何对科学数据进行统计、编目和制定目录等。最后,针对科学数据管理的馆员,需明确工作制度、工作分工、职责范围等。

5.2 探索服务模式,嵌入科学数据生命周期

科研活动中产生的科学数据“生成—捕获—组织—评估—分析—存储—再利用”的动态循环为数据生命周期^[31]。面向科学数据管理,须以科研生命周期理论为基石,嵌入数据生命周期,对照不同阶段,制定对

应阶段的服务模型与框架并探索服务模式。

典型的数据生命周期有UKDA研究数据生命周期模型、Data ONE数据生命周期模型等^[32],结合科研主体的发展现状和特点,尽管不同生命周期模型的适用范围和对象各有侧重,但其功能均囊括了为机构和科研人员制定数据管理计划、数据获取、数据描述、数据保存、数据共享与再利用等共同的阶段要素。可借鉴或发展现有生命周期模型全流程,采用发放调查问卷或座谈等方式明确科研人员的科学数据管理需求,充分调研设计科学数据管理计划模板并辅助科研人员制定计划,在制定数据汇交政策的前提下为科研人员提供可操作的数据汇交相关培训,同时设计并明确科学数据的元数据描述模板,生成元数据记录,在此基础上形成科学数据分级分类目录和索引,此后在充分调研科研人员对数据安全和数据权益看法的前提下,为科学数据的复用制定协议与规则,完成科学数据的全流程管理。

5.3 提升专业能力,打造科学数据管理团队

负责机构知识库建设和维护的图书馆馆员(管理员)应当积极转变角色,结合科学数据管理的生命周期,提高科学数据管理服务的意识和技能,形成具备专业素养的科学数据管理服务团队。我国尚未对负责机构知识库的馆员工作进行明确的划分和区别,但当前欧美高校图书馆已开始致力于培养包含数据管理咨询人员(data counselor)、科学数据馆员(research data librarian)、数据验证人员(data verification)、数据分析人员(data analyst)、数据可视化分析人员(data visualization analyst)、数据存储馆员(repository librarian)、数据管理专家(data specialist)和数据服务协调人才(data service coordinator)等科学数据管理人才^[33],如牛津大学典藏库ORA由4名主管和7名元数据助理组成服务团队^[34]。我国也应当从宏观上调整和扩展机构知识库的服务,在系统开发层面,提升技术能力;在数据服务层面,制定并执行科学数据管理工作规范;在系统互操作层面,需要具备整合和管理来自其他科研管理系统的科研人员、科研项目、各阶段科学数据和科研机构信息的能力。此外,除了对机构知识库中现有的科学数据进行审核、修改和统计,在实现科学数据获取与管理的基础上,还可通过适当调整管理员权限,培训各科研院系科研秘书并对其开放数据权限,实

现各科研管理相关系统与机构知识库的数据管理和共享,打通图书馆馆员与科研人员之间联系的桥梁,开展机构知识库在科学数据管理服务方面的需求调查,有针对性地服务科研人员和团队。

5.4 强化互联合作,打通科学数据管理链条

馆员应基于开放框架设计,从技术层面上提升机构知识库与机构内外其他系统的合作与交流,打通科学数据传播和共享的阻碍。

中国高校机构知识库联盟(CHAIR)促进了我国高校机构知识库建设初期的发展,同时也为各高校建设机构知识库提供了交流和沟通的平台,尽管联盟高校间尚未真正实现数据的开放和共享,但仍可借助联盟这一渠道加强与其他机构知识库的互联。除加入联盟之外,馆员可采用开源技术框架促成机构知识库与其他科学数据管理平台的互联:①基于数据开放接口实现机构知识库与其他系统(如专门的科学数据管理平台、机构内部科研管理系统等)的互操作,使得第三方系统可以完成机构知识库科学数据的检索与集成,西北工业大学知识管理系统采用的便是该技术方案,北京大学2015年上线的北京大学开放研究数据平台与北京大学机构知识库完成了互操作,平台的数据集通过机构知识库进行保存与服务^[18];②对开源的DSpace架构机构知识库基于CRIS技术二次开发,融合科研管理架构,如香港大学机构知识库;③对科研管理系统进行机构知识库模块开发,欧美一些机构知识库便是如此。此外,还可基于如Sitemap的自动生成和发布接口支持机构知识库数据被全球搜索引擎索引和发现,提高知识库内容的公共传播和发现影响,提高资源影响力。

6 结语

科学数据管理服务需要依托于信息庞大、功能齐全的平台,而机构知识库逐渐重视科学数据发现、管理与复用,致力于拓展数据增值服务。两者相辅相成,既可为科学数据管理提供平台与服务,又能支持机构知识库的可持续发展,使后者更好地为科研服务。为此还需在优化政策引导、制定数据管理规范、组建专业人才队伍以及革新互联技术方面做持续性的探索与努力。

参考文献

- [1] 黄鼎成. 科学数据共享管理研究 [M]. 北京: 中国科学技术出版社, 2002: 33-34.
- [2] 涂志芳, 刘兹恒. 国内外科学数据管理服务评价研究与实践进展 [J]. 图书馆建设, 2021 (2): 108-117.
- [3] Next Generation Repositories: Behaviours and Technical Recommendations of COAR Next Generation Repositories Working Group [EB/OL]. [2022-07-28]. <https://www.coar-repositories.org/files/NGR-Final-Formatted-Report-cc.pdf>.
- [4] 邢文明. 我国科研数据管理与共享政策保障研究 [D]. 武汉: 武汉大学, 2014.
- [5] 张玉娥, 王永珍. 欧盟科研数据管理与开放获取政策及其启示——以“欧盟地平线2020”计划为例 [J]. 图书情报工作, 2017, 61 (13): 70-76.
- [6] 周晓燕, 宰冰欣. 澳大利亚高校科研数据管理政策制定研究 [J]. 图书馆建设, 2017 (2): 63-70.
- [7] 黄国彬, 屈亚杰. 英国科研资助机构的科学数据共享政策调研 [J]. 图书馆论坛, 2017, 37 (5): 124-132.
- [8] 魏悦, 刘桂锋. 基于数据生命周期的国外高校科学数据管理与共享政策分析 [J]. 情报杂志, 2017, 36 (5): 153-158.
- [9] 袁晓明, 王美琴. 基于微服务架构的Globus科研数据管理平台分析 [J]. 数字图书馆论坛, 2021 (12): 22-27.
- [10] 汪满容, 刘桂锋, 刘琼. 新加坡高校图书馆科研数据管理服务调研与启示 [J]. 图书馆学研究, 2018 (9): 22, 64-71.
- [11] 王继娜. 国外高校图书馆科学数据管理服务的调研与思考 [J]. 情报理论与实践, 2019, 42 (8): 159-167.
- [12] 张贵香, 刘桂锋, 梁炜. 我国科研数据管理理论与服务研究进展述评 [J]. 情报理论与实践, 2020, 43 (6): 187-193.
- [13] 陈和, 陈小慧. 2019年机构知识库发展报告 [EB/OL]. [2022-07-05]. <https://dspace.xmu.edu.cn/handle/2288/174250>.
- [14] WILSON J A J, MARTINEZ-URIBE L, FRASER M A, et al. An institutional approach to developing research data management infrastructure [J]. International Journal of Digital Curation, 2011, 6 (2): 274-287.
- [15] KIM Y, KIM S. Institutional, motivational, and resource factors influencing healthscientists' data-sharing behaviours [J]. Journal of Scholarly Publishing, 2015, 46 (4): 366-389.
- [16] 陈和, 林静. 2000—2018年国内机构知识库研究可视化分析 [J]. 知识管理论坛, 2019 (5): 2-8.
- [17] 吴建中. 从开放获取到“超越开放获取”——“下一代机构知识库”给予我们的启示 [J]. 国家图书馆学刊, 2018, 27 (6): 3-8.

- [18] 崔海媛, 孙超, 罗鹏程. 新一代机构知识库的关键技术和发展趋势研究[J]. 图书情报工作, 2019, 63(19): 96-104.
- [19] 都平平, 彭琳, 李雨珂, 等. 下一代高校机构知识库中的灰色科研数据资源规划管理与复用策略研究[J]. 兰台世界, 2021(8): 109-114.
- [20] 王思敏, 阮楠, 吴丹. 面向科研数据管理的精准型机构库建设与实践[J]. 图书馆学研究, 2019(2): 58-64.
- [21] 梁娜, 张晓林. 机构知识库的互操作需求和互操作规范框架[J]. 现代图书情报技术, 2013(9): 1-7.
- [22] Trinity's Access to Research Archive [EB/OL]. [2022-08-01]. <http://www.tara.tcd.ie/>.
- [23] 肖可以. 基于满足科研信息管理需求的CRIS应用平台研究[J]. 图书馆学研究, 2015(1): 41-47.
- [24] 孙惠英, 朱莲花, 陈淑英, 等. 日本大学图书馆与日本国立情报研究所合作建设数字资源的经验与启示[J]. 数字图书馆论坛, 2021(4): 25-30.
- [25] 从学术典藏(IR)到当前科研信息系统(CRIS)——如何和为何[EB/OL]. [2022-07-26]. <https://www.docin.com/p-1572543327.html>.
- [26] 国务院办公厅. 科学数据管理办法[EB/OL]. [2022-07-26]. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.
- [27] 董成立. 谈高校科研管理及其信息管理系统[J]. 科技管理研究, 2009, 29(5): 274-276.
- [28] 唐玫, 吉洪湖, 胡娅萍. 超声速飞行器综合热管理系统优化设计[J]. 推进技术, 2022, 43(1): 50-60.
- [29] Open Doar [EB/OL]. [2022-07-26]. <https://v2.sherpa.ac.uk/opensoar/>.
- [30] 邓支青. 高校机构知识库内容质量控制策略研究[J]. 情报杂志, 2018, 37(9): 173-178.
- [31] INMON W H. Data architecture: a primer for the data scientist [J]. Data Architecture A Primer for the Data Scientist, 2014 (13-15): 57-62.
- [32] 张宇, 刘文云. 基于数据生命周期的高校机构知识库科研数据服务研究[J]. 图书馆学研究, 2021(3): 71-80.
- [33] 崔涛, 徐建国. 图书馆视角下科学数据管理人才培养模式研究[J]. 图书馆建设, 2017(3): 84-89.
- [34] Contact ORA [EB/OL]. [2022-07-28]. https://libguides.bodleian.ox.ac.uk/ora/contact_ORA.

作者简介

唐玫, 女, 1992年生, 硕士, 馆员, 研究方向: 学科信息服务与科学数据服务, E-mail: may_0315@nwpu.edu.cn。

赵婉忻, 女, 1984年生, 硕士, 副研究馆员, 研究方向: 信息计量与科学评价。

李晶, 女, 1980年生, 硕士, 馆员, 研究方向: 数字图书馆开源软件等。

师俏梅, 女, 1970年生, 副研究馆员, 研究方向: 情报分析与信息服务。

Development Situation and Enlightenment of the Institutional Repositories for Research Data Management

TANG Mei ZHAO WanXin LI Jing SHI QiaoMei

(Northwestern Polytechnical University Library, Xi'an 710072, P. R. China)

Abstract: Under the background of open science, scientific activities are becoming more intensive, and research data management is becoming more and more important. Based on the basic functions of data warehousing, traditional institutional repositories is gradually expanding to research data management and knowledge value-added services. Based on the investigation and discussion of the current development situation of the construction of institutional repositories at home and abroad, key technologies are discussed, taking the construction practice of Knowledge Management System in Northwestern Polytechnical University as an example, the service framework of institutional repositories for research data management is constructed, and finally suggestions of the institutional repositories construction for research data management is proposed: optimizing management policies, formulating research data management norms, embedding research data life cycle, building professional research data management teams, and strengthening the interconnection and cooperation between platforms.

Keywords: Open Science; Research Data Management; Institutional Repositories

(收稿日期: 2022-08-22)