

我国科学数据共享平台建设及服务内容研究

王翠萍 宋雯琪 姜鑫妍

(东北师范大学信息科学与技术学院, 吉林长春 130117)

摘要: 为了解我国科学数据共享平台建设现状及服务内容, 探索各科学数据共享平台间服务方式的差异, 对比各平台建设中存在的问题。利用科学数据共享平台官方网站, 从建设背景、组织规范、服务内容和能力调查分析我国14个典型的科学数据共享平台的总体情况。研究发现, 我国科学数据共享平台重视数据汇交服务, 数据汇交流程较为完善, 平台数据影响力不断提升, 数据共享服务内容较为丰富, 但仍存在平台组织规范缺少绩效评价内容、数据与科技成果间关联性不足以及数据利用和共建共享程度不足的问题。因此, 建议科学数据共享平台需进一步完善元数据标准、数据分类标准, 形成统一的引用标准, 提供统一的集成检索服务, 并注重与其他平台间的关联建设, 提高科学数据的共享利用水平。

关键词: 科学数据; 科学数据共享平台; 科学数据中心; 数据服务; 平台建设

DOI: 10.3772/j.issn.1674-1544.2023.03.002

CSTR: 15994.14.issn.1674.1544.2023.03.002

中图分类号: G350

文献标识码: A

Research on the Construction and Service Content of Scientific Data Sharing Platform in China

WANG Cuiping, SONG Wenqi, JIANG Xinyan

(School of Information Science and Technology, Northeast Normal University, Changchun 130117)

Abstract: In order to understand the current situation and service content of scientific data sharing platform construction in China, explore the differences of service modes among various scientific data sharing platforms, and compare the problems in the construction of various platforms. This paper analyzes the overall situation of 14 typical scientific data sharing platforms in China through visiting the official website of the scientific data sharing platforms and investigating the construction background, organizational standards, service content and service capacity. It is found that the scientific data sharing platform attaches importance to data collection and exchange services, the data collection and exchange process is relatively complete, the platform's data influence continues to increase, and the content of data sharing services is relatively rich. However, there are problems such as the lack of performance evaluation content in the platform organization specification, the insufficient correlation between data and scientific and technological achievements and lack of data utilization, co-construction and sharing. Therefore, the scientific data sharing platform needs to further improve metadata standards and data classification standards, form a unified reference standard, provide a unified integrated search service and pay attention to the association construction with other platforms to improve the sharing

作者简介: 王翠萍 (1974—), 女, 东北师范大学信息科学与技术学院教授, 博士生导师, 博士, 研究方向为信息组织与服务、信息分析与咨询; 宋雯琪 (1998—), 女, 东北师范大学信息科学与技术学院硕士生, 研究方向为信息资源开发与管理 (通信作者); 姜鑫妍 (1999—), 女, 东北师范大学信息科学与技术学院硕士生, 研究方向为信息资源管理研究。

收稿时间: 2022年10月25日。

and utilization level of scientific data.

Keywords: scientific data, scientific data sharing platform, scientific data center, data service, platform construction

0 引言

科学数据是重要的科技资源,是科学研究的基础。在开放科学背景下,世界各国逐渐意识到科学数据的价值。科学数据共享平台可以提供科学数据的管理、共享及分析挖掘等服务,有利于促进数据有效管理和利用。国际组织、政府部门和科研机构都积极参与科学共享平台的建设,目前已建成诸如re3data、科学数据银行Science DB等在世界范围有较大影响力的科学数据存储库。随着科学数据共享平台建设的实践不断发展,近年来国内外关于科学数据平台的研究取得了一定的研究成果。国外研究主题主要集中在某一数据共享平台的功能服务介绍^[1-3]及价值^[4]等方面,并且有研究关注到了区块链这类新兴技术在科学数据共享平台中的应用^[5-6]。国内关于科学数据共享平台领域的相关研究主要集中在国家科技基础条件平台建设及发展^[7-9]、绩效评价^[10]、国外数据中心建设经验启示^[11-13]、各数据中心建设服务总结^[14-17]等方面。除此之外,有研究通过调查国家级科学数据中心的管理模式、资源体系、功能体系,总结经验提出科学数据中心建设的框架^[1]。现有研究主要集中于调查或介绍总结某一数据共享平台的建设及服务现状,虽然也有研究关注到了国内外科学数据管理平台建设实践的差异^[19-21],但是对我国不同数据共享平台间建设现状及服务效果的对比研究较少。

随着科学数据开放共享建设进程的推进,我国投入了大量资金及资源开展科学数据共享平台建设,国内科学数据共享平台数量不断增长。我国科学数据共享平台按建设责任主体可分为四大类,即国家级数据中心、省级数据平台、高校科研院所数据中心以及企业建设的科学数据平台。目前,我国国家科技基础条件平台已建成了20个数据中心及31个资源库,收录了自然及

人文社科领域的观测、实验及调查等多种类型数据集。此外,平台中心还对其收录的科学数据提供科技资源标识服务,以支持和完善科技资源的管理与使用。在逐一访问每个国家级数据中心官方网站,大致了解其在平台建设、数据描述、共享以及引用的总体情况的基础上,本文选取国家科技基础条件平台中具有代表性的10家国家数据共享平台^[22]以及广东省科技资源共享服务平台^[23]、CnOpenData^[24]、北京大学开放研究数据平台^[25]以及西南财经大学中国家庭金融调查与研究中心^[26]4个较为典型的其他类型数据共享平台。其中,CnOpenData是商业公司开发的科学数据共享平台。本文对这14个数据共享平台的官方网站进行网络调查,从组织规范、服务内容和服务能力等方面分析我国科学数据共享平台建设的优势及不足,以期为我国数据共享平台的建设提供一定的参考。

1 科学数据共享平台建设现状

1.1 科学数据共享平台建设背景

在科学数据开放共享的实践历程中,我国的科学数据开放建设经历了漫长的发展阶段。科技部于2001年启动了气象科学数据共享试点作为我国第一个科学数据共享平台,随后又于2003年设立了科技基础条件平台建设专项,由此开始了我国科学数据共享工程的建设。构建高水平的科学数据共享体系有利于解决我国科学数据资源分散重复、封闭垄断等问题,是国家创新体系的重要组成部分。国家科技基础条件平台门户应用系统自2006年开始部署,由国家科技资源共享服务工程技术研究中心提供技术支持,建成了国家科技基础条件平台中心中国科技资源共享网这一科技资源在线服务系统,并于2007年举行了揭牌仪式,于2009年9月正式开通运行。国家科技基础条件平台建设通过运用现代信息技术,

以新型管理机制和运行模式,对全社会科技基础条件资源进行了重新整合,以促进数据资源的共享和利用。

全社会的科技资源拥有者均可使用中国科技资源共享网公布科技资源信息。通过中国科技资源共享网站,用户可以直接访问与其关联的20个国家科学数据中心及31个国家生物种质与实验材料资源库。同时,网站还汇总不同所属平台的数据资源,形成资源目录,方便用户进行检索查询。国家科技基础条件平台中心承担着平台建设发展战略、运行管理、标准规范建设、国际合作与宣传、培训以及考核评估和监督等职责。目前,共享网在国家层次上,搭建了全国范围的逻辑统一、高度集成、高效共享的科技资源网络服务体系,有效推动了科技资源的统筹管理和共享服务,实现了以信息共享带动实物共享的平台理念。

省级科学数据共享平台由科技资源共享网发展而来,并逐渐向机构知识库和智库方向转变。调查发现:四川省、湖北省科技信息资源共享平台主要资源类型都以文献为主,科学数据资源较少;山东省社会科学中心包括人才库、成果库及数据中心则更偏向于机构知识库性质;省级数据平台更侧重于建设政府开放数据,部分省级科学数据平台出现无法访问的问题;广东省科技资源共享服务平台从2010年发展至今已成为科学数据及相关资源管理的综合性平台,实现了对省内科技资源的有效整合。

2018年,国务院办公厅印发的《科学数据管理办法》指出有关科研院所、高等院校和企业等法人单位是科学数据管理的责任主体^[27]。高校科研院所为责任主体建设的科学数据共享平台可大致分为两类:一是以北京大学开放研究数据平台为代表的综合类科学数据平台。二是专题类数据平台,如复旦大学社会科学数据平台等。西南财经大学中国家庭金融调查与研究是专题类调查数据共享平台,可通过申请获取数据。商业化的科学数据共享平台更为常见,CnOpenData是涵盖各行业领域的数据平台,已有180个数据资源

库,并处于不断更新中,包括自有数据、公共数据和第三方数据,数据资源较为丰富。

1.2 科学数据共享平台组织规范

为了解我国科学数据共享平台的建设情况,调查包括了平台组织单位和组织规范,详见表1。由表1可知,国家数据平台大部分是由中国科学院建设和运行,基本上在平台网站都会有相应的组织规范,通过制定章程、制度或管理规范明确数据中心机构运行的任务、用户的职责和科学数据共享权利和义务。如国家高能物理科学中心通过用户委员会章程、科学数据合作共享制度、数据共享联盟章程和账号管理制度形成平台管理规范,多方面详细规定平台建设宗旨与用户应遵守的行为规范。为尊重知识产权,保障数据安全,保护数据所有者和访问者的权利,国家青藏高原科学数据中心和国家人口健康科学数据中心的官方网站公布了隐私政策与免责声明。国家海洋科学数据中心数据在权限说明中规定了普通用户、个人认证用户和单位认证用户的每日最高数据下载量,其平台管理办法包括了组织职责、数据资源建设、汇交及共享、考核评价、经费使用等内容。省级科学数据平台一般由本地区科学技术厅负责,并接受国家科技基础条件平台中心的业务指导。广东省科技资源共享服务平台管理规范包含了国家标准法规及地方推荐标准,有利于保障科技资源的交流共享。而企业科学数据平台及高校科学数据平台在组织规范方面,仅有北京大学开放研究数据平台在用户指南栏目中发布了使用手册和使用条款,另外调查的两个平台网站皆无相应的组织规范。通过各平台所公布的组织规范可以看出,在运行规范及管理方面,虽然国家级和省级数据共享平台网站基本都有相关制度规范,但主要是针对运行管理、数据管理及共享,缺少绩效考核评价的内容。而企业和高校专题类数据共享平台在对其网站组织规范方面建设不足,缺乏相关的运营管理细则。

调研各科学数据共享平台中的标准规范板块所公示的内容发现,各平台在管理规范及政策法规方面的侧重差异,除国家对地观测科学数据

表1 科学数据共享平台建设基本情况

序号	平台单位	组织单位		组织规范
		依托单位	主管部门	
1	国家高能物理科学数据中心	中国科学院高能物理研究所	中国科学院	关于管理规范有《高能物理数据中心用户委员会章程》《高能物理科学数据合作共享制度》等 关于技术规范有《高能物理科学数据汇交管理办法》《论文标注规范》《先导专项江门中微子实验数据管理办法(试行版)》等
2	国家空间科学数据中心	中国科学院国家空间科学中心	中国科学院	章程有《国家空间科学数据中心章程》 关于标准规范有CDF、FITS、VOTable、HDF5、SAO.XML
3	国家对地观测科学数据中心	中国科学院遥感与数字地球研究所	中国科学院	
4	国家青藏高原科学数据中心	中国科学院青藏高原研究所	中国科学院	隐私政策; 使用条款和免责声明
5	国家生态科学数据中心	中国科学院地理科学与资源研究所	中国科学院	隐私政策; 数据管理与共享标准规范有观测指标和观测技术规范、科技资源描述规范、信息化规范、资源整合操作流程规范、数据共享规范
6	国家冰川冻土沙漠科学数据中心	中国科学院寒区旱区环境与工程研究所	中国科学院	关于标准规范有规章制度《国特殊环境与灾害研究网络章程》《中国特殊环境与灾害研究网络绩效考核评估办法》等和业务规范《地面气象观测规范(中国气象局编制)》等
7	国家地球系统科学数据中心	中国科学院地理科学与资源研究所	中国科学院	标准规范
8	国家人口健康科学数据中心	中国医学科学院	卫生健康委	隐私与声明有PHDA 免责声明、PHDA 隐私和安全政策
9	国家地震科学数据中心	中国地震台网中心	地震局	关于标准规范有数据共享规章制度《地震科学数据共享管理办法》、数据共享标准规范《地震科学数据 数据发布规范》等以及地震相关法规《中华人民共和国防震减灾法》等
10	国家海洋科学数据中心	国家海洋信息中心	自然资源部	《国家海洋科学数据共享服务平台管理暂行办法》; 数据权限说明有普通用户、个人认证用户、单位认证用户
11	广东省科技资源共享服务平台	广东省科学技术厅	广东省科学技术厅	相关法规有《中华人民共和国科学技术进步法》等; 相关办法、规范有《科学数据管理办法》等; 关于国家标准有《科技计划形成的科学数据汇交 技术与管理规范》(GBT 39912-2021)等; 广东省地方标准有《大型科学仪器设施共享服务平台数据交换规范》(DB44/T 2192-2019)等
12	CnOpenData	河南汉数信息技术有限公司		
13	北京大学开放研究数据平台	北京大学图书馆	北京大学	用户使用手册、用户使用条款
14	中国家庭金融调查与研究中心	西南财经大学中国家庭金融调查与研究中心	西南财经大学	

中心、国家地球系统科学数据中心无法在其网站首页查询到标准规范相关内容及文件外, 国家高能物理科学中心公布的标准规范侧重于平台运营自身的管理制度和章程, 明确了平台组织机构、数据管理办法及数据汇交流程等内容; 国家空间科学数据中心在章程中介绍了平台的使命和任务、组织管理与职责及运行机制等内容, 同时

提供了针对不同数据类型的数据标准规范; 国家青藏高原科学数据中心和国家人口健康科学数据中心从尊重知识产权和维护网络安全的角度, 向数据用户提出了平台数据使用条款和数据使用免责声明; 国家生态科学数据中心提出了其数据管理与共享标准规范体系框架, 介绍了收集、整理、加工、存储、服务及信息化方面标准, 但具

体制度及标准内容无法从其网站获取; 国家冰川冻土沙漠科学数据中心根据不同来源及不同类型数据制定了更具针对性的规章制度和业务规范, 并针对各类资源制定了考核指标与评估方式以评价平台的运行服务绩效; 国家地震科学数据中心所提供的标准规范主要分为数据共享规章制度、数据共享标准规范、地震相关法规 3 个部分的内容; 国家海洋科学数据中心数据权限说明规定了普通用户、个人认证用户和单位认证用户的每日最高数据下载量, 其平台管理办法包括了组织职责、数据资源建设、汇交及共享、考核评价、经费使用等内容。广东省科技资源共享服务平台管理规范涵盖数据汇交、元数据以及平台管理等内容, 并且重视数据安全问题。北京大学开放研究数据平台用户使用手册主要针对用户使用数据行为以及分享数据行为两个方面, 用户使用条款包括隐私政策、免责声明、著作权内容, 并对用户注册、提交以及数据引用下载作出规定, 注重保护数据分享者及利用者双方的权利, 形成良好的科学数据共建共享生态环境。

分析上述科学数据共享平台所提供标准规范可以看出, 各平台通过制定章程、制度或管理规范来明确数据中心机构运行的任务、用户的职责和科学数据共享权利和义务。在运行规范及管理方面, 各数据共享平台网站在制度规范的数量及侧重方向上虽有所差异, 但主要集中于对平台运行管理、平台数据管理及共享的相关标准等内容, 除此之外就是对平台数据用户的使用行为进行限制。但仅有少部分平台的规范涉及平台绩效考核, 各平台的标准规范中还应考虑平台的资源数量与质量、资源服务数量、资源服务成效等指标的考核及评估, 从更客观的角度对平台的建设、运行和服务效果进行规范。

2 科学数据共享平台服务分析

2.1 科学数据共享平台数据服务内容

在科技计划项目实施过程中, 同步产生海量科学数据资源, 是我国科学数据的主要积累途径^[28]。根据《科学数据管理办法》和《国家科技

资源共享服务平台管理办法》要求, 各级科技计划管理部门应建立规范化的科学数据汇交及管理机制, 在科学数据汇交完成后再验收科技计划项目。平台数据汇交流程一般包括用户注册、(元)数据汇交方案或计划、数据审查和存档, 数据汇交需签订数据使用协议以保护知识产权和共享数据, 促进数据重用。为保证数据汇交工作的顺利进行, 数据中心会举办涵盖数据汇交的要求、流程、系统使用及其他注意事项内容的培训活动。数据共享平台提供的软件工具主要有数据可视化工具、模型库、统计分析工具等。在对所调研的 14 个科学数据共享平台提供的数据服务进行汇总后, 形成了如表 2 所示的调查表。从表 2 可以看出: 所调研的 10 家国家科技基础条件平台数据中心均为政府预算资金资助的基金、专项项目及社会资金项目产生的科学数据提供数据汇交支持; 广东省科技资源共享服务平台为平台基地及科技基础条件建设类项目形成的科技资源提供数据汇交服务; 其他由非政府部门主导建设的科学数据共享平台的主要服务内容则主要集中于平台数据库的开放下载方面, 并未涉及数据汇交的内容。

中国科技资源共享网的资源目录提供了下属各科学数据共享平台的网址链接, 通过此链接可以快速访问不同所属平台中的科学数据, 各科学数据共享平台内部以统一的格式对相关数据信息进行介绍。具体字段包括数据资源的基本信息、所有者信息、共享方式、资源提交机构信息、资源服务机构信息、标识信息(包括类目信息、类目代码、分类标准等)、标识注册机构信息等, 并可以通过链接直接访问发布地址。但是, 不同数据库所提供的科学数据其描述方式有所不同。在具体访问上述科学数据中心对其科学数据的描述和公开情况后, 发现部分科学数据库不仅提供了多种数据标识方式, 而且对学者引用其中数据的方式进行了明确的规范和限定。有些描述则较为简单, 只是将其中科学数据内容进行开放共享, 并未对其使用方式有所规范。任何机构或个人只有在申请后才可以访问数据集内的具体数

表2 科学数据共享平台数据服务内容调查

序号	平台名称	数据服务内容		
		软件工具	数据检索	服务方式
1	高能物理科学数据中心	实验软件、通用软件	关键词检索、CSTR标识检索	数据访问、专业软件、共享与分析服务
2	国家空间科学数据中心	统计、数值、同化模型、其他工具	普通检索、高级检索	数据共享服务、专业软件工具服务和论文数据贮藏库服务、数据深度对接、定制化专题服务
3	国家对地观测科学数据中心	卫星产品	关键词检索、数据集检索、数据产品检索	提供在线下载、离线下下载、在线分析3种服务方式
4	国家青藏高原科学数据中心	大数据分析方法库、模型库	普通检索、数据导航	在线共享为主、离线共享为辅的数据共享服务
5	国家生态科学数据中心	分析平台、可视化平台	普通检索、筛选条件	数据获取服务、实物科技资源服务
6	国家冰川冻土科学数据中心	模型库	普通检索、筛选条件	提供数据共享、数据定制、数据挖掘及在线计算等多种服务
7	国家地球系统科学数据中心	—	普通检索、筛选条件	在线服务、离线服务、委托服务、主动服务、应急服务、其他服务等多种服务方式
8	国家人口健康科学数据中心	数据采集、处理、分析、可视化、存储、管理、检索、查询工具	普通检索、高级检索	数据注册服务、汇交服务、长期保存服务、定制化服务、数据接口服务、专题服务、科普服务、应用示范等多种服务
9	国家地震科学数据中心	—	全站检索	数据下载、电话咨询、信息咨询以及根据用户需求在线下订单制作专题数据产品等服务
10	国家海洋科学数据中心	数据可视化	全站检索	数据共享、专题信息产品等服务
11	广东省科技资源共享服务平台	—	关键词检索、筛选条件	科技资源检索、科学数据汇交、仪器信息报送、省内仪器查重、创新平台评估、科技专题服务等
12	CnOpenData	—	按数据库进行分类无检索条件	数据最新动态、期刊最新动态、数据下载
13	北京大学开放研究数据平台	—	普通检索、高级检索	引用和下载、数据分享和共建、查询数据使用统计报告
14	中国家庭金融调查与研究中心	—	普通检索	数据下载

据，只有在注册账号后才能上传和下载相关科学数据。此外，各平台还为用户提供了数据分析软件等操作工具。综合考虑各科学数据中心的建设情况，发现各中心的服务板块、内容和模式均存在差异。在数据的分类及类目设置方面，各平台根据不同学科领域的科学数据特点分类，缺乏统一的分类规范。在数据描述方面，不同平台的元数据描述标准不同，大多数数据中心所提供的科学数据给定了基于国家标准的CSTR科学数据标识符编码，也有数据中心提供了科学数据的DOI

编码等其他数据标识编码。在科学数据引用规范方面，有的科学数据中心提供了学科数据引用说明和致谢格式，规范了后续科研人员引用数据集的格式。

广东省科技资源共享服务平台接受国家科技基础条件平台中心的业务指导，除运行模式相似外，在对平台数据管理方面也引入CSTR科学数据标识符编码，但在数据获取方面，此平台内包含的不同数据库资源设有不同的共享途径和共享范围，数据资源无法直接获取，且无相关的数据

引用规范。CnOpenData、北京大学开放研究数据平台和中国家庭金融调查与研究的服务内容主要围绕平台数据的开放共享，用户在注册账号并通过审核后即可对平台数据进行下载，但在数据描述方面，CnOpenData和中国家庭金融调查与研究对其内部数据库信息描述较为简便，只有北京大学开放研究数据平台对其数据提供了DOI编码并制定了这个数据集的引文著录格式。

2.2 科学数据共享平台数据服务能力

数据和数据集总量、数据下载量以及平台访问量是反映一个数据平台科学数据质量的直接指标。从表3平台数据使用数量和服务效果的调查数据可以发现，就科学数据共享平台服务数量而言，大多数国家科技基础条件平台下属的科学数据中心都以公开的方式提供了其数据集总量、数据下载量、数据访问量以及平台访问量的统计数据，所提供的数据资源数量及种类均较为丰富。但广东省科技资源共享服务平台、CnOpenData、北京大学开放研究数据平台和中国家庭金融调查与研究4个科学数据共享平台的数据使用情况及数据下载量等指标公开程度较差，甚至有些平台并未公开此类数据。但从已公开其服务案例

及访问数据的几家平台来看，各平台间的数据集数量和平台访问人次数量差异明显，由于不同科学数据平台所侧重的学科主题间差距较大，且都集中于其领域内的专业主题，因此不同平台所面对的不同使用者特征及其使用偏好也是导致不同平台之间使用效率存在差异的重要原因。

从调研结果看，各科学数据共享平台均提供了数据检索功能，包括基本检索和高级检索等多种方式，可以同时检索本平台及分数据中心的数据资源，方便用户查找平台内部的数据资源。国家科技基础条件平台建成了多级数据分类组织模式，通过整合下属各数据中心资源，在保障数据资源持续更新的基础上，加强了面向专题服务的跨领域、跨学科的数据资源深度挖掘和集成。在数据关联方面，各平台实现了科学数据的内部关联，但不同平台之间的关联仅仅是网站链接方式。虽然平台提供数据下载和使用信息等，但却无法直接对接科研成果或科技文献等内容，在科学数据与科技产出间关联建设方面较为薄弱。广东省科技资源共享服务平台的服务对象主要面向广东省内各级政府部门、事业单位和大型企业，平台数据对个人用户的开放性不足，且内部数据

表3 科学数据共享平台服务调查

序号	平台名称	数据集使用量			平台服务效果		
		在线数据集/个	数据量	数据下载量	平台访问量/人次	服务案例数量/件	分中心数量/个
1	高能物理科学数据中心	—	3.761 PB	—	—	13	4
2	国家空间科学数据中心	327	336.2 TB	182.3 TB	203 092	65	5
3	国家对地观测科学数据中心	3 943	2.5 PB	552 万条	2 074.51 万	8*	14
4	国家青藏高原科学数据中心	5 421	222 TB	—	21 131 848	84	—
5	国家生态科学数据中心	703	—	—	—	4	3
6	国家冰川冻土科学数据中心	—	32.3 TB	2.4 PB	171 923 912	10	7
7	国家地球系统科学数据中心	—	—	—	—	11*	9
8	国家人口健康科学数据中心	16 170	1.27 PB	—	3 414 383	10	—
9	国家地震科学数据中心	103	260 TB	147 TB	128 万	1*	10
10	国家海洋科学数据中心	—	—	—	1 607 059	9	8
11	广东省科技资源共享服务平台	1 363	—	—	—	41	7
12	CnOpenData	180	—	—	—	—	—
13	北京大学开放研究数据平台	355	—	1 068 390	—	—	—
14	中国家庭金融调查与研究	3	—	—	—	—	—

注：标*代表在官方网站未明确提及服务案例数量，但提供专题服务内容。调研数据源自各数据中心官网，调查时间截至2022年8月。

尚未形成与其他平台间的关联。CnOpenData、北京大学开放研究数据平台和中国家庭金融调查与研究中心数据主要以调查统计类数据为主,数据类型较少且服务方式也较为单一,使用效果相较于前两种由政府主导建设的共享平台有较大差距。

国家青藏高原科学数据中心是服务案例最多的科学数据中心,并且是国内首个通过Nature数据期刊Scientific Data认证的数据仓储中心,这有利于国内数据中心国际影响力的提升。其国际服务案例——中国植被功能型图(1 km)服务于科学研究,下载及引用次数较高,成效显著。国家科学数据中心是每年举办的“共享杯”竞赛的数据提供平台,如人口健康科学数据中心为比赛提供人口健康科学数据资源,地震科学数据中心为比赛提供地震领域相关科学数据资源。除了典型服务案例可以表明科学中心的服务效能外,提供专题服务也是数据共享平台的特色服务之一。国家对地观测科学数据中心提供地理空间开源软件专题服务,为科学研究服务。然而,软件数量多但访问量和下载量偏低,用户利用率较低。在一定程度上与开源软件的性质有关,开源软件的下载安装不一定需要通过科学数据共享平台,其他下载途径和渠道更为普遍,因此也出现了平台软件利用率较低的现象。

3 科学数据共享平台服务建设优势与不足

3.1 优势分析

3.1.1 数据汇交流程较为完善

目前,我国已建立了国家政府资金资助项目的科学数据管理和汇交制度,颁布的国家标准和政策文件对数据汇交内容和数据汇交流程提出了明确的规定。在此背景下,国家科技基础条件平台中心全部提供数据汇交服务,并且在官方网站的栏目分类中设置了单独板块。国家青藏高原科学数据中心展示了详细的数据汇交流程、步骤及注意事项。国家人口健康科学数据中心针对数据汇交开展了专门的培训活动,在官网上可以找到项目数据汇交培训的视频。数据汇交培训视频包括了数据汇交背景知识、方案介绍、数据汇交

计划的制定和审核、数据仓储项目数据汇交功能演示等内容,而且近期还举办了数据汇交培训讲座。由此可见,我国科学数据共享平台十分重视数据汇交服务,这既是国家政策的要求也是数据共享平台建设的重点内容。

3.1.2 数据影响力范围较为广泛

科学数据影响力是指科学数据的应用在学术、社会、政策等多方面产生的正向影响,一般包括学术影响力和社会影响力两种^[29]。国家科技基础条件平台建设致力于打破“数据孤岛”,服务于科学研究、社会民生等。在技术和软件工具方面,数据中心提供了模型库和数据可视化工具,为科研人员提供更为专业和便捷的服务。在科学研究和公共服务上,典型案例有国家空间科学数据中心月球与行星探测数据的在线公众信息服务,帮助公众了解、学习和使用嫦娥月球数据,网站累计访问量达17万余人次,提升了嫦娥月球数据的影响力和社会经济价值。国家青藏高原科学数据中心所有数据都是中英文标注,有利于国内外科学研究的交流合作。随着科学数据平台建设的发展,服务成效明显提升,数据影响力不断扩大。

3.1.3 数据共享服务内容较为丰富

整体而言,我国科学数据共享平台建设内容主要围绕着科学数据的组织、描述、存储、共享等方面,意在整合科技资源,为用户提供数据服务。根据上述调研的各科学数据共享平台访问情况,可以发现现有平台可以在线浏览网站内部数据库资源。为保障数据的精确获取,各平台均已具备检索功能,提供元数据检索服务,部分平台还可以提供在线或离线咨询及数据分析等服务。除此之外,在数据库内容建设和更新方面,数据共享平台逐渐从建设单一领域的数据库资源库转变为建设涵盖更大规模、更多类型的集成数据库,形成了多学科、多领域的各类专题数据库。同时,在科学数据的开放获取方面,上述调研的共享平台数据可以直接在官方网站下载,无操作门槛,便利程度较强。在数据集的使用浏览情况、下载管理方面也提供了相应的关联记录,数

据开放程度较高。

3.2 不足分析

3.2.1 平台组织规范的绩效评价有待加强

对 14 个科学数据共享平台的组织规范调查发现, 目前只有国家冰川冻土沙漠科学数据中心和国家海洋科学数据中心官方网站提及绩效考核评价内容。《国家海洋科学数据共享服务平台管理暂行办法》明确了考核内容、方式、时间和评价等级, 重点考核平台科技资源整合能力、服务成效、组织运行管理及专项经费使用情况等内容。国家冰川冻土沙漠科学数据中心颁布了绩效考核评估办法和细则, 对考核目标、内容、评价方式等作出了说明, 详细规定了考核指标和计分方法, 值得其他数据中心借鉴。绩效评价对于科学数据共享平台的运行管理和服务具有重要作用, 平台组织规范内容应考虑绩效评价。当然, 绩效考核是一项系统工程, 制定考核评价指标体系时应考虑到各平台数据资源及服务的特点, 体现了科学性和客观性, 有效激励科学数据共享平台的发展。

3.2.2 平台数据与科技成果间关联建设不足

科学数据与科技文献关联可以解决科学数据分散的问题, 促进科学数据的流转和利用^[30]。就目前科学数据平台的建设现状而言, 虽然各数据中心所提供的科学数据体量大, 但数据类别却因受到学科的限制性而仅集中于某一领域。在元数据描述格式方面, 各平台间也缺乏统一的标准规范。在数据资源整合方面, 各平台多集中于对其内部数据进行组织与管理, 对其他平台的资源仅能提供平台链接而无法直接进行关联与整合。在平台科学数据的标准引用方面, 各平台针对其数据引用提供了引用格式与引用声明, 然而对数据本身的描述信息较少, 并且缺乏一致的格式标准。在对科技文献中的科学数据引用行为进行溯源时, 难以实现精确的定位与识别, 导致科学数据与科技成果之间的关联情况不明确。对科学数据影响力的评估及数据重用效率的提升造成一定的困难, 对保障数据生产者的知识产权造成一定的障碍。

3.2.3 平台数据利用和共建共享程度不高

各平台对于科学数据共享权限、共享范围和共享方式的限制情况可以反映这个平台的数据开放共享程度。我国科学数据共享平台大多对其内部数据实行分类分级管理, 依据不同数据级别控制数据共享范围。这方法虽然能有效防止平台数据被泄露和滥用, 但也限制了用户对其数据的使用。在平台数据资源共建共享建设方面, 国家级和省级科学数据共享平台虽然都设有数据汇交服务板块, 但却未形成良好的平台间数据交换机制和数据共建模式; 高校科学数据共享平台如北大数据平台暂时只支持本校的认证用户创建和发布数据, 仅在团队内部设立了数据共建机制; 企业所主导的科学数据共享平台与其他平台的合作较少, 主要是通过吸纳第三方数据库资源的方式扩充其内部数据。目前, 我国各科学数据共享平台尚未将更多数据整合纳入统一的数据共享交换体系, 无法实现有效的跨平台、跨领域的数据共享流动和供需对接。

4 结论与建议

科学数据管理与共享工作的规范化是一个从基本理念到法律制度, 再到法律实现的过程^[31]。科学数据相关政策的制定与不断完善是推动科学数据开放共享的有效政策支撑和制度保障。对上述 14 个科学数据共享平台建设现状的调查可以发现, 在平台的组织规范建设方面, 所调研的 14 个平台中仅有 6 个平台公示了较为完善的数据管理及共享规范。但在制度规范的数量及侧重方向上有所差异, 仅有少部分平台的规范涉及平台的绩效考核, 因此各平台内部仍需对组织管理政策进行完善, 形成覆盖平台管理、用户管理、数据管理等全方面的管理制度。在对科学数据的组织与分类建设方面, 在所调研的 14 个平台中, 仅有国家生态科学数据中心、国家人口健康科学数据中心和国家地震科学数据中心 3 个平台公布了内部元数据字段的描述及编写规范, 而在实际建设过程中, 各平台的元数据标准、数据分类标准等并无统一的格式标准, 且各平台间差异较大。

这不仅影响数据组织的效果,而且出现用户在引用平台数据时著录格式规范性不足等问题,对科学数据生产者的版权保护和对科研成果的溯源验证造成困难。为解决这一问题,在目前的用户指南与数据共享政策的基础上,科学数据共享平台还需完善元数据标准、数据分类标准等内容,同时结合不同学科、不同领域的的数据特征形成统一的引用标准。

在数据服务与科研成果关联方面,所调研的科学数据平台都能够通过开放获取数据资源的方式为科学项目提供帮助,但很少有平台能够进一步提供数据加工分析或定制服务。在数据汇交建设方面,所调研的10家国家科技基础条件平台数据中心和1个省级科技资源共享服务平台均具备此功能,对科学数据的收录及利用方面重视程度较高,但由企业和高校所承办的3个科学数据共享平台尚未开通数据汇交服务,对科学数据的整合集成方面建设不足。在科学数据发布方面,大多数平台仅仅将相关数据集在该平台内部进行共享,通过链接形式访问其他平台的数据集,无法直接对接有关科技成果与科技文献。为提高科学数据的利用效率,科学数据中心应注重与其他平台间的关联建设,将数据集和数据平台、数据资源等进行链接,以关联数据形式发布科学数据及科技论文。同时,还应完善科学数据平台功能,聚合更多的数据库资源,提供统一的集成检索服务。不仅科学数据中心具备数据存储功能,而且用户可以利用数据开放共享平台实现学术间的交流与共享。

参考文献

- [1] KEERTHANA I P, ABAHAI T A. DSPAA: a data sharing platform with automated annotation[J]. *Procedia computer science*, 2016, 6(8): 45-53.
- [2] DOEL T, SHAKIR D I, PRATT R, et al. GIFT-Cloud: a data sharing and collaboration platform for medical imaging research[J]. *Computer methods and programs in biomedicine*, 2017, 32(2): 181-190.
- [3] D'ANCA A, CONTE L, NASSISI P, et al. A multi-service data management platform for scientific oceanographic products[J]. *Natural hazards and earth system sciences*, 2017 (17): 171-184.
- [4] WYSEL M, BAKER D, BILLINGSLEY W. Data sharing platforms: how value is created from agricultural data[J]. *Agricultural systems*, 2021, 45(8): 1-12.
- [5] VISARA U, VIKAS J, ARNO A, et al. LUCE: a block-chain-based data sharing platform for monitoring data license accountability and Compliance[J]. *Blockchain: research and applications*, 2022, 4(3): 1-10.
- [6] KUMI S, LOMOTEY R K, DETERS R. A Blockchain-based platform for data management and sharing[J]. *Procedia computer science*, 2022, 11(8): 95-102.
- [7] 朱亮, 孟宪学, 赵瑞雪, 等. 国家农业科学数据共享中心资源建设探析[J]. *数字图书馆论坛*, 2017(11): 15-20.
- [8] 邓仲华, 黄雅婷. “互联网+”环境下我国科学数据共享平台发展研究[J]. *情报理论与实践*, 2017, 40(2): 128-132.
- [9] 李正超. 国内科学数据共享平台建设现状及发展策略研究[J]. *图书馆理论与实践*, 2018(8): 108-112.
- [10] 陈丽娜, 司海平, 方洸, 等. 国家科技基础条件平台服务绩效评价研究[J]. *科技管理研究*, 2016, 36(18): 262-266.
- [11] 袁红卫, 黄松, 刘嫣. 麻省理工学院科学数据管理与共享平台调研及启示[J]. *图书馆学研究*, 2019(13): 95-101, 82.
- [12] 石蕾, 高孟绪, 徐波, 等. 欧美建设发展科学数据中心的经验及对我国的启示[J]. *中国科技资源导刊*, 2022, 54(3): 31-36, 110.
- [13] 吴思竹, 王安然, 修晓蕾, 等. 欧美生物医学科学数据中心建设及启示[J]. *数字图书馆论坛*, 2022(4): 2-10.
- [14] 张华吉, 张晓宇. 国家人口健康科学数据共享平台药学数据中心建设工作思考[J]. *科研信息化技术与应用*, 2018, 9(4): 55-59.
- [15] 杨雅萍, 姜侯, 孙九林. 科学数据共享实践: 以国家地球系统科学数据中心为例[J]. *地球信息科学学报*, 2020, 22(6): 1358-1369.
- [16] 张连翀, 李国庆, 李静. 国家对地观测科学数据中心建设及服务[J]. *卫星应用*, 2021(7): 8-14.
- [17] 张华吉, 张晓宇. 国家人口健康科学数据共享平台药学数据中心建设工作思考[J]. *科研信息化技术与应用*, 2018, 9(4): 55-59.
- [18] 刘杨, 陈晓玲, 全志薇, 等. 国家级科学数据中心建设对吉林省科学数据中心建设的示范效应[J]. *中国管理信息化*, 2022, 25(3): 173-178.

(下转第67页)

- casting Innovation Pathways (FIP) for new and emerging science and technologies[J]. Technological forecasting & social change, 2013, 80 (2): 267-285.
- [2] LORI P. What's new in the 2022 gartner hypecycle for emerging technologies[EB/OL].[2022-08-10]. <https://www.gartner.com/en/articles/what-s-new-in-the-2022-gartner-hype-cycle-for-emerging-technologies>.
- [3] Department of Energy Office of Science. Top 10 emerging technologies of 2021[EB/OL].[2021-11-16]. https://www3.veforum.org/docs/WEF_Top_10_Emerging_Technologies_of_2021.pdf.
- [4] Accelerate innovations in emerging technologies[EB/OL].[2023-02-17]. https://science.osti.gov/grants/Lab-Announcements/-/media/grants/pdf/lab-announcements/2023/LAB_23-3010.pdf.
- [5] 俞阳. 欧盟未来新兴技术计划概况及进展[J]. 全球科技经济瞭望, 2014, 29(4): 1-6.
- [6] 李仕明, 肖磊, 萧延高. 新兴技术管理研究综述[J]. 管理科学学报, 2007, 10(6): 12-12.
- [7] 杨春霞, 罗小安, 侯宏飞, 等. 美国国家科学基金会设施的管理及启示[J]. 中国科学院院刊, 2015(3): 347-353.
- [8] DAY G S, SCHOEMAKER P, GUNTHER R E. Wharton on managing emerging technologies[M]. Hoboken, New Jersey: John Wiley & Sons, Inc., 2000: 1-435.
- [9] ROTOLO D, HICKS D, MARTIN B R. What is an emerging technology?[J]. Research policy, 2015, 44(10): 1827-1843.
- [10] 鲁若愚, 张红琪. 基于快变市场的新兴技术产品更新策略研究[J]. 管理学报, 2005(3): 317-320.
- [11] 赵洪江, 陈学华, 苏晓波. 新兴技术、新技术、高技术及高新技术概念辨析[J]. 企业技术开发, 2005(11): 40-62.
- [12] 王田苗, 郝雨飞, 杨兴帮, 等. 软体机器人: 结构、驱动、传感与控制[J]. 机械工程学报, 2017, 53(13): 1-13.
- [13] RUS D, TOLLER M T. Design, fabrication and control of soft robots[J]. Nature, 2015, 521(7553): 467-475.
- [14] WANG Y P, YANG X B, CHEN Y F, et al. A bio robotic adhesive disc for underwater hitchhiking inspired by the remora suckerfish[J]. Science robotics, 2017, 2(10): 1-9.
- [15] UMRao S, TABASSIAN R, KIM J, et al. MXene artificial muscles based on ionically cross-linked $Ti_3C_2T_x$ electrode for kinetic soft robotics[J]. Science robotics, 2019, 4(33): 1-11.
- [16] TANG Y C, CHI Y D, SUN J F, et al. Leveraging elastic instabilities for amplified performance: spine-inspired high-speed and high-force soft robots[J]. Science advances, 2020, 6(19): 1-12.

(上接第18页)

- [19] 卫军朝, 张春芳. 国内外科学数据管理平台比较研究[J]. 图书情报知识, 2017(5): 7-107.
- [20] 李舸, 柏永青, 王卷乐, 等. 中美地球系统科学数据共享平台对比分析[J]. 中国科技资源导刊, 2018, 50(2): 6-13.
- [21] 崔旭, 赵希梅, 王铮, 等. 我国科学数据管理平台建设成就、缺失、对策及趋势分析: 基于国内外比较视角[J]. 图书情报工作, 2019, 63(9): 21-30.
- [22] 中国科技资源共享网[EB/OL]. [2022-08-30]. <https://www.escience.org.cn/>.
- [23] 广东省科技资源共享平台[EB/OL]. [2022-12-23]. <https://www.gdkjzy.net/#/index>.
- [24] 北京大学开放研究数据平台[EB/OL]. [2022-12-23]. <https://opendata.pku.edu.cn/>.
- [25] 中国家庭金融调查与研究中心[EB/OL]. [2022-12-23]. <https://chfs.swufe.edu.cn>.
- [26] cnopendata[EB/OL].[2022-12-23]. <https://www.cnopendata.com/>.
- [27] 中国政府网. 国务院办公厅关于印发科学数据管理办法的通知[EB/OL].[2022-12-23]. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.
- [28] 徐波, 王瑞丹, 王卷乐, 等. 科技计划项目科学数据汇交共性机制研究[J]. 中国科技资源导刊, 2021, 53(1): 9-14.
- [29] 王毅萍, 马建玲. 国外科学数据影响力研究进展[J]. 图书情报工作, 2017, 61(7): 118-126.
- [30] 崔旭, 赵希梅, 王铮, 等. 我国科学数据管理平台建设成就、缺失、对策及趋势分析: 基于国内外比较视角[J]. 图书情报工作, 2019, 63(9): 21-30.
- [31] 秦顺, 邢文明. 开放·共享·安全: 我国科学数据共享进入新时代: 对《科学数据管理办法》的解读[J]. 图书馆, 2019(6): 36-42.