

Mendeley Data平台开放科学数据 实践及启示*

卢焱 王鸢飞 刘洪冰 袁雪 陈斯翰 刘敏娟
(中国农业科学院农业信息研究所, 北京 100081)

摘要: 高效的研究数据集成发现和管理工具推动研究数据的开放共享。本文以Elsevier公司Mendeley Data科研数据管理与共享平台为研究对象, 从数据治理流程、集成的开放生态系统、数据存储管理和共享功能方面进行调研, 分析、总结该平台开放数据实践中围绕数据标准在流程设计和系统开发方面的优势, 为我国高校图书馆、研究型图书馆以及信息服务机构开展开放数据资源建设和研究数据服务提供参考借鉴。

关键词: 研究数据; 开放数据; 数据管理; 数据共享; Mendeley Data

中图分类号: G203.1 DOI: 10.3772/j.issn.1673-2286.2021.09.003

引文格式: 卢焱, 王鸢飞, 刘洪冰, 等. Mendeley Data平台开放科学数据实践及启示[J]. 数字图书馆论坛, 2021(9): 19-26.

1 集成发现和管理工具推动研究数据开放共享

科学体系正在经历重大的转变, 从以研究人员为主导的专业体系, 转变为有多个利益相关者的开放创新体系^[1]。在开放科学中, 开放数据作为重要要素, 其实践促进科研协作和透明度, 并且提升科研成果的重现性和影响力^[2-3]。近年来无论是研究人员、图书馆、资助者、政策制定者, 还是出版方、各类倡议组织、资源开发和策展方, 各利益相关方都对推动研究数据的有效共享给予了更高关注。研究人员对数据共享持积极态度, 根据莱顿大学的调查结果^[1], 73%的受访者表示已获得已发表的研究数据将有利于自身研究, 69%的受访者表示共享研究数据对于他们所在领域的研究很重要, 同时有64%的受访者愿意让别人访问自己的研究数据。2020年STM出版商协会开始实施研究数据实践, 将2020年定为“STM研究数据年”, 启动了STM研究数据项目。根据该项目调查^[4], 2020年参与该项目的出版商中, 采用数据政策的期刊从29%增加到52%, 而包

含数据可用性声明(DASs)的文章数量增加一倍多。

然而研究数据的分享实践面临很多挑战, 根据全球7 000余位科研人员反馈^[5], 46%的科研人员认为将数据进行共享的最主要障碍是不知道如何将数据按照可展示、有用的方式进行有效的组织, 其他困扰因素包括对版权和授权条款的不了解(37%)、不知道在何处使用何种数据仓储(33%)、缺乏时间(26%)和担心成本(19%)。因此研究数据共享并不应该仅是对开放数据政策开展研究和制定宏观数据管理计划, 在推进数据管理实践层面也同样重要。如何为科研人员提供良好的数据管理支撑服务, 创建便捷的数据共享路径, 帮助研究人员更容易访问和使用数据, 也应当引起重视。

数据共享的主要目标是使研究数据以标准化的方式被其他研究人员重用。FORCE11工作组于2016年发布了《FAIR数据原则》, 代表了数据共享和重用的黄金标准^[6]。开放数据不仅要可发现可访问, 更需要能够被理解、可信赖和可重用。将数据存储标准化的数据仓储中可以增加公开和重用的机会, 目前许多组织提供了用于归档数据集的存储库^[7]。随着研究数据的日益丰

* 本研究得到中国农业科学院科技创新工程专项“国家农业科技文献数字资源建设与服务”(编号: CAAS-ASTIP-2021CX025-AII) 资助。

富, 高效的研究数据集成发现和管理工具越来越不可或缺。本文选取Mendeley Data平台进行调研, 它是Elsevier公司2015年开发的科研数据管理与共享平台, 本文针对其数据治理流程、开放关联特征、数据存储管理和共享功能进行分析, 总结其开放数据实践过程中, 遵循FAIR标准实施流程和系统设计的优势经验, 为图书馆开展开

放数据资源建设和研究数据服务带来启示。

2 Mendeley Data的研究数据治理流程

Mendeley Data的数据治理流程包括数据收割、数据管理发布、数据存档和提供访问4个环节(见图1)。

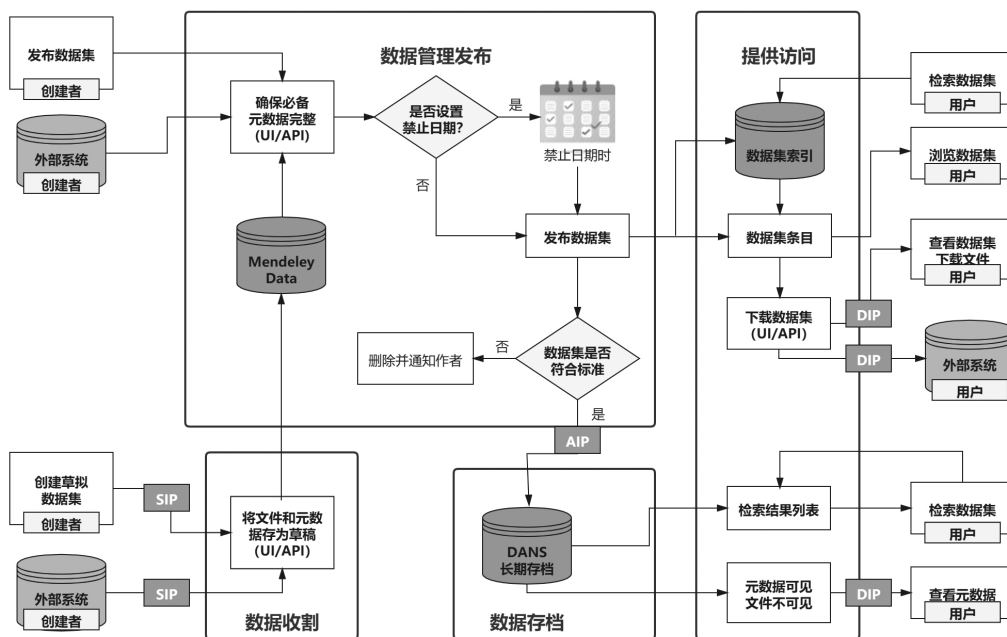


图1 Mendeley Data基于开放档案信息系统(OAIS)模型的数据治理流程^[8]

2.1 数据收割

Mendeley Data索引的科研数据一方面来自于个人用户通过网页界面创建数据集; 另一方面来自于外部系统通过配置公共API创建数据集, 如Elsevier的文章发表投稿系统EVISe。个人用户或者外部系统需要按照规范元数据(见表1)对数据集进行描述。

2.2 数据管理发布

数据集通过UI和API进行发布, 以便公众进行访问。在数据集发布前, 系统将从形式上确保必备元数据的完整性。数据集可立即发布, 也可设置禁止日期延迟其发布, 若设置禁止日期的情况下发布数据集, 仅其标题是立即公开的, 其余的元数据和文件直至禁止日期才公开。公开时, 数据集将从DataCite接收一个新的数字对象标识符(Digital Object Identifier, DOI)。数据集发布后, 由Mendeley Data平台审核员从内容上进行

表1 Mendeley Data元数据

名称	说明	约束
Title	标题	必备
Contributors	姓名、ORCID iD、电子邮箱、所属机构、贡献	必备
Institutions	与本研究有关的机构(如大学)及/或单位(如系/学院)	
Categories	来源于Elsevier's OmniScience taxonomy	必备
Description	关于数据的文本描述	必备
Steps to reproduce	关于重复该数据集对应研究步骤的文本描述	
Related links	与该数据集相关的任何数据集、文章、软件或其他实体的链接	
License	版权许可	必备

审核, 确保其为具有科学性质的研究数据, 尤其是要与包含研究结果的研究论文进行区别, 并且确保不涉及以下情况: 已经公开发布过, 具有DOI; 含有可执行文档但缺乏独立、详细的文件描述; 包含有版权的内容(音

频、视频、图像等);包含敏感信息等。而不符合审核标准的数据集将被删除并通知作者。

2.3 数据存档

数据集依照上述标准被审核为有效后,随即被归档到数据归档和网络服务(Data Archiving and Network Services, DANS)长期资源库^[9]中。Mendeley已经与DANS签署协议,确保所有已发布和有效的数据集都将被永久存档,并且所有存档的数据集将独立于Mendeley Data站点,而在DANS中可用。通过DANS存档,保证了所有格式文件的长期保存和DANS优选数据格式文件的长期可用性。

2.4 提供访问

科研用户可以从Mendeley Data网站查看和下载数据集。外部系统也可以通过Mendeley Data公共API查询和下载任何已发布的有效数据集。平台提供元数据和文件的全文搜索功能,并且可按数据类型、来源仓储、学科类别过滤搜索结果。同时所有数据集都分配有永久DOI,也能够通过已发表的相关论文或其他相关数据集进行链接访问。

3 Mendeley Data形成了关联集成的开放数据生态系统

Mendeley Data通过与DANS^[9]、OpenAIRE^[10]、DataCite^[11]、Scholix^[12]等一系列科研数据基础设施系统进行集成,促进科研数据在全球范围内的开放关联。在Elsevier内部与ScienceDirect、Scopus、Scival等产品进行关联,将科研数据管理共享全面整合入其学术研究数字化服务生态。

(1) 利用DANS解决数据长期保存问题。DANS是一个长期存档的供应机构,隶属荷兰皇家科学院KNAW和荷兰科学研究基金会NWO^[9]。Mendeley Data与DANS达成协议对所有有效的已发布数据集进行永久存档,协议将确保数据集分配的DOIs将始终被解析为一个包含可用数据集元数据和文件的web页面。

(2) 与欧洲科研开放获取基础设施OpenAIRE(Open Access Infrastructure for Research in Europe)及其Zenodo数据仓储创建索引。OpenAIRE^[10]作为欧

盟的研究门户网站,是欧洲最大的科研数据发布中心,通过建立索引,当数据集在Mendeley Data平台发布时,也将同步发布到OpenAIRE。

(3) 通过DataCite元数据索引与数千个科研数据仓储创建索引,从而索引千万级别的数据集。DataCite是一家于2009年在英国成立的国际非营利性组织,由多家机构联合发起,为科学数据创建元数据集^[11],在Mendeley Data平台发布的数据集,元数据也将发送至DataCite。同时DataCite作为科学数据标识符注册中心,为研究数据和其他研究产出分配DOI,增强数据搜索能力,建立数据共享生态和高效引用机制。

(4) 通过Scholix链接研究论文和科研数据。学术链接交换(Scholix)是一项倡议,为其成员提供一个全面的互操作性框架,通过收集和交换研究数据与文献之间的链接,创建开放的全球信息生态系统^[12]。其成员需使用统一的信息模型来表示数据和文献之间的链接,以及标准的交换格式和协议来实现信息交换。Mendeley Data作为Scholix的成员之一,通过将其元数据(包括相关文献的链接)发送到DataCite来为Scholix做出贡献,DataCite是Scholix用于聚合数据集和文献之间链接的信息集散中心之一,从而使链接被广泛地发现。

(5) 与Elsevier内部产品关联,将科研数据服务纳入其学术研究数字化服务生态。Mendeley Data数据集通过Scholix与Scopus数据库中相应出版物进行关联访问,进一步借助Scopus数据库中的规范信息对数据集作者、所属机构等元数据进行增强;在ScienceDirect以及SSRN文献出版流程中,Mendeley Data则被用于提交或公开与文献相关的数据集;研究数据和文献一样也可通过计量指标度量其影响力,Mendeley Data复用了Elsevier的PlumX指标体系,包括下载、查看和数据引用(源自DataCite和Crossref)等,以及替代计量学指标来反映社交媒体的关注度和对媒体、政策的影响力。

4 Mendeley Data的功能

Mendeley Data作为覆盖科研数据全生命周期的模块化研究数据管理云平台,具有数据检索、数据存储和管理以及数据监控功能,各功能模块可独立使用,也可以与机构现有的解决方案集成在一起以确保科研管理工作的连续性。

4.1 数据检索

Mendeley Data可作为研究数据开放搜索引擎进行数据检索,其索引了全球2 000多个公共研究数据仓储中超过2 810万个数据集,也包括Mendeley Data管理平台的数据集,并且将各个来源的研究数据进行元数据和数据文件的深度标引,由此提供字段的高级检索,可对检索结果数据文件内容进行在线预览,同时突出显示搜索词,以使用户快速评估和选择检索记录。

4.2 数据存储和管理

Mendeley Data作为数据存储解决方案,使机构能够存储文本、表格、图像、统计数据、数据库、置标语言等各种类型的研究数据^[13],数据通过云服务器进行托管也可以选择存储在本地服务器。数据库采用标准且丰富的元数据对数据记录进行管理,并允许机构用户自定义元数据。同时支持大规模的数据协作,使研究人员能够与机构内外部合作者在同一个项目环境中对研究数据进行组织、注释和共享。此外,还提供一系列高级数据管理工具,如版本控制和审核 workflow 等,促进研

究数据管理的最佳实践。

4.3 数据监控

数据监控模块能追踪检索全球2 000多个数据库中超过2 810万个数据集,机构用户能借助其索引自动追踪监控本机构科研数据产出情况。Mendeley Data从2 000多个通用和领域数据仓储收集索引研究数据,按照OpenAIRE元数据方案进行规范描述,经数据清洗删除无效索引记录,再利用自然语言处理和文本挖掘技术,综合DataCite、Scholix、Scopus等多个来源的信息,通过添加出版物链接、作者标识符(ORCID、Scopus Author ID等)和机构标识符(Scopus Affiliation ID等)的方式来丰富数据集元数据,从而形成底层数据搜索语料库(见图2)。数据监控模块在此基础上,比对作者信息和机构隶属信息的元数据,确定数据集的归属,即使在原始元数据中缺少归属信息的情况下也能依据增强的元数据进行判断。因此使得机构科研管理者、图书馆员能够追踪所在机构研究人员发布的数据集,进而进行监控和审核,并且无论这些数据集是发布于机构数据存储库中,还是发布于外部的通用或领域数据仓储中。

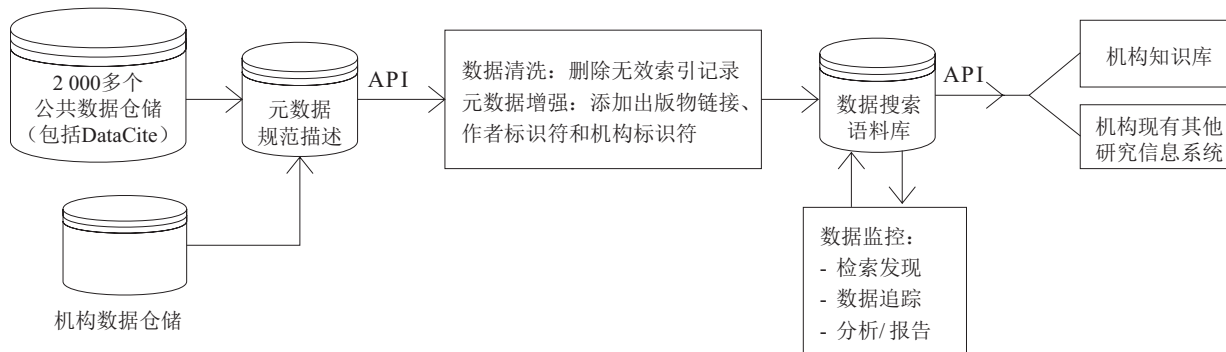


图2 数据监控模块的工作原理

此外,数据监控模块通过建立机构用户专门的研究数据门户,集中展示本机构的科研数据产出情况,提升机构数据的可发现性和可复用能力。数据监控模块还提供API服务将数据集元数据集成到机构知识库(IR)和机构现有研究信息系统(CRIS)中,实现自动匹配作者和链接到出版物。此外,还支持导出数据报表和生成报告,以进行批量审核和分析。

5 Mendeley Data的数据发布和共享机制

用户利用平台进行科研数据发布和分享,但科研数

据的所有权和控制权应当由用户掌握而非平台。Mendeley Data平台由用户根据需求对数据集进行“私密”或“公开”两种状态的权限设置,从而控制数据集的发布。私密状态意味着数据集尚未发布,用户可在该状态下对数据集进行编辑和删除。用户设置权限选项为“公开”状态则意味着对数据集进行发布,类似于正式发表一篇文章,数据集一旦发布则不能进行编辑和删除,平台将公开数据集的名称、所有者、版本、发布时间以及DOI。而即便是发布数据集,用户仍可以通过设置截止日期来延迟公开的时间,这对研究者而言有充分的自主选择权确保在论文发表之前让研究数据处于非公开

状态, 又能保证编辑和审稿人可见。此外, 在用户发布科研数据时, 平台针对各种数据类型, 提供了15种资料公开授权许可方式 (见表2) 供其选择, 包括常见的一系列CC授权条款、软件许可条款 (MIT、Apache、BSD、

GPL) 以及硬件许可条款, 其中CC0许可相对最不受约束, Mendeley Data平台发送至DataCite的元数据使用该许可方式。

表2 Mendeley Data提供的15种资料公开授权条款

授权条款	数据类型	说明
CC0 1.0	所有数据类型	可对该数据集进行复制、修改、分发和执行, 即使是出于商业目的, 上述操作均无须请求许可
CC BY 4.0	所有数据类型	提供适当信用信息的条件下, 数据集可以被分享、复制和修改。可提供指向CC BY许可的链接, 并明确指出存在修改之处, 或者其他方式证明版权持有人支持对数据集的使用。 此外, 数据集中标识为属于第三方的任何内容可能需要进一步的权限
CC BY NC 3.0	所有数据类型	提供适当属性的条件下, 可对该资料进行自由地改编、复制或重新分发, 并且不将该资料用于商业目的
MIT	软件	本版权声明和许可声明应包括在软件的所有副本或实质性部分
Apache-2.0	软件	属于永久的、全球的、非排他性的、免费的、免版税的、不可撤销的版权许可, 以复制、派生、公开展示、公开演示、转授权和分发作品或衍生作品的源形式或其他对象形式
BSD 3-clause	软件	重新发布源代码必须保留原有版权声明、条件列表和免责声明。 以二进制形式重新发布必须在发布文件和/或提供的其他材料中重申原有版权声明、条件列表和免责声明。 未经事先书面许可, 不得使用版权所有人或其贡献者的姓名来支持或推广该软件衍生的产品
BSD 2-clause	软件	重新发布源代码必须保留原有版权声明、条件列表和免责声明。 以二进制形式重新发布必须在发布文件和/或提供的其他材料中重申原有版权声明、条件列表和免责声明
GPLv3	软件	通用公共许可是针对软件和其他类型的作品的免费、非营利版权许可
GPLv2	软件	通用公共许可是针对软件和其他类型的作品的免费、非营利版权许可
LGPL	软件	GNU宽通用公共许可 (GNU Lesser General Public License) 允许他人使用并对软件进行集成, 且无须发布自身组件的源代码
MPL-2.0	软件	Mozilla公共许可2.0版本是一种简单的非营利许可证, 具有“文件级”的公共版权, 旨在鼓励贡献者共享其对源代码所做的修改, 并且仍然允许其受最小限制的与其他授权代码进行结合
CeCILL	软件	CeCILL许可授予用户自由复制、修改和发布许可软件的权利, 对软件的修改必须根据CeCILL发布
CeCILL-B	软件	CeCILL-B许可授予用户修改和重新发布软件的自由, 但其强制规定了对原件的归属义务 (通过引用原始记录来实施)
CERN OHL	硬件	CERN开放硬件许可证 (CERN OHL) 管理硬件设计文档的使用、复制、修改和分发, 以及相关产品的制造和分发
TAPR OHL	硬件	TAPR开放硬件许可证 (OHL) 为硬件项目提供了一个框架, 与用于开放源码软件的框架类似, TAPR允许任何人使用OHL作为其硬件项目的许可证, 但从未更改形式的使用为前提

6 Mendeley Data的实践遵循FAIR原则

开放高效的研究数据管理和共享工具一定是遵循FAIR原则的^[4], 通过分析Mendeley Data的数据治理流程、其集成的开放生态系统, 并进一步梳理其功能, 对Mendeley Data研究数据管理流程和系统设计中使研究数据具有可查找性、可访问性、互操作性和可重用性特征的解决方案进行总结, 以供开放科学数据资源建设进行借鉴。

(1) 可查找性。Mendeley Data数据仓储中所有数据集都将被分配DOI, 包括数据集下级文件夹和文件,

甚至是数据集的不同版本, 并且通过数据集元数据注册和索引, 使其能在多个研究数据资源发现系统中被检索, 除Mendeley Data平台本身提供的检索之外, 在Google Dataset Search、DataCite、OpenAIRE、Open Science Framework都可以被发现。

(2) 可访问性。发布的数据集通过标准化通信协议进行访问, 借助DOI标识符通过HTTPS协议用浏览器进行访问或者REST API访问。在访问过程中允许在必要时进行身份验证和授权, 机构用户可以集成与行业标准SAML 2.0协议兼容 (如Shibboleth) 的机构认证。即便数据集未来不再可用, 元数据仍可被访问。

Mendeley Data系统支持“墓碑”DOIs机制,即使数据集从系统中删除,DOI仍能解析数据集页面。同时由于在DANS长期存档,确保了对数据资源的永久访问,若Mendeley Data平台停止服务,数据集将通过DOIs解析并指向存储在DANS中的副本,数据集和元数据将永远可用。

(3) 互操作性。Mendeley Data使用标准的、广泛应用的、可访问的方式来表示知识,其采用JSON格式作为数据交换语言,通过HTTPS/REST协议同时也支持OAI-PMH元数据互操作协议,与其他研究数据管理工具集成进行数据收割和推送;通过语义链接关联研究论文、软件、其他数据集等研究对象;支持Dublin Core和schema.org标准元数据框架,提供标准元数据字段的标识符(作者及其隶属机构字段)和受控词汇表(学科类别和许可条款字段)。

(4) 可重用性。Mendeley Data基于丰富精确的元数据描述不仅提供数据溯源信息,也促进数据重用。通过许可条款受控词汇表提供清晰和可访问的数据使用许可;同时鼓励用户编辑“Steps to reproduce”字段,对重演该数据集对应研究步骤和方法进行文字描述;还允许机构和研究人员添加自定义元数据模板来进一步丰富元数据,尤其是可以添加某些特定领域的元数据字段,并且可将自定义元数据字段取值范围设置为现有的分类法,以便于发现和重用。

7 对研究数据资源建设和科研数据服务的启示

Mendeley Data的研究数据实践充分证明了它不仅是研究数据存储工具,更是工作工具,依据数据重用和共享标准为研究人员提供存储和发布的工作环境,将研究人员作为共享和重用数据实践的核心使之真正受益,也通过这种自下而上的方法缩小政策与开放数据基层实践之间的差距,为开展研究数据资源建设和科研数据服务提供可借鉴的实际措施。

7.1 加强研究数据资源的元数据建设

元数据的价值一直是科学数据各项研究和实践中探讨的重要问题和关键问题,元数据标准对科学数据的发现、共享和再利用至关重要^[15-16]。机构内部自建研究数据仓储应使用统一元数据描述标准,提高数据的

互操作性。而图书馆或信息服务机构进行开放科学数据资源建设时,可通过建立统一元数据仓储,对分散的开放科学数据资源进行集成汇聚,以便集中发现,除对公共数据仓储的数据集进行索引之外,商业出版社也掌握丰富的研究数据资源,在2020年STM研究数据项目的驱动下,有21家出版商13 064种学术期刊参与该项目,并且研究数据联盟(Research Data Alliance, RDA)的数据政策标准化和实施利益小组为所有期刊和出版商开发了研究数据政策标准框架^[17],从出版商的角度积极推动研究数据共享,因此加强与出版商在研究数据元数据方面的合作也十分必要。除此之外,从Mendeley Data的实践还可以看出,对元数据进行深度标引,配合统一分类系统和受控词表等知识组织工具的使用,将有利于资源的深度揭示。

7.2 积极加入研究数据国际联盟和倡议组织

从Mendeley Data可以看出,Elsevier公司作为全球领先的出版企业在研究数据领域非常活跃,是DataCite、Scholix、FORCE11、Research Data Alliance等多个联盟和组织的成员,在研究数据资源汇聚、关联、分享、利用,以及标准和指导原则制定等各方面都有布局。国内科研机构,尤其是国家级的科技信息服务机构也应当为开放数据的未来提前做准备,积极加入各类研究数据国际联盟和倡议组织,与各类学术生态系统参与者(如资助方、机构、出版商和政策制定者)开展广泛的合作。借助DataCite这类社区的数据资源集成、数据注册等机制获得更大范围的研究数据共享,目前DataCite在中国有5家会员,分别是北京大学、清华大学、中国国家基因库(China National GeneBank)、中国散裂中子源(China Spallation Neutron Source, IHEP)和全国地质资料馆(National Geological Archives of China);同时充分利用全球研究数据基础设施^[18],减少技术成本投入;参与研究数据相关标准和指导原则的制定,并遵循通用的标准和原则开展具体实施。

7.3 拓展研究数据服务类型

围绕研究数据服务有很多系统的研究,都指出应推行面向科研全生命周期的嵌入式服务,包括发现访问、数据存储和管理、培训教育等^[19-23]。Mendeley Data的解决方案使一些服务更具可操作性,如数据链接服

务, 指导作者在数据存储库与在线文章之间创建双向链接来实现论文和数据的更容易被发现和访问。此外以标准化且深度标引的元数据为基础, 也可进一步拓展对科研数据利用价值和潜力的开发, 使科研数据和文摘数据一样, 能够通过数据挖掘和情报分析, 为不同类型的用户提供数据驱动的科研管理决策支撑服务。例如: 通过追踪本机构研究人员发布的数据集情况, 更好地响应资助基金要求, 辅助管理部门制定机构相关的数据管理政策; 跟踪全球、区域或机构研究数据产出, 洞察学科发展趋势和热点前沿等^[24]。

7.4 推进研究数据引用和人员识别实践应用, 完善激励和评估机制

数据共享的障碍之一是研究人员需要付出额外的时间和工作, 却无法获得直接回报。对研究数据的引用可以改变这种情况, 通过引用提高发布者的学术声誉, 以此激励其分享行为。数据引用是支持数据重用的学术生态系统的一部分, FORCE11工作组为数据引用制定了一套指导原则^[25], 并为出版商和学术团体提出了实施建议^[26]。与论文产出成果的评估类似, 引用指标和替代计量指标, 同样适用于研究数据的影响力评估, 因此未来数据引用及相关评价指标也可纳入当前基于论文引用的评估体系和激励机制。而推行这方面应用的重要前提是对数据集发布者的身份识别及其归属机构的认定, Mendeley Data通过元数据增强进行归属认定的做法值得借鉴。

参考文献

- [1] MEIJER I, BERGHMANS S, COUSIJN H, et al. Open Data: the researcher perspective-survey and case studies [R/OL]. [2021-07-01]. https://www.elsevier.com/_data/assets/pdf_file/0004/281920/Open-data-report.pdf.
- [2] U.K. Government. Open Data White Paper: unleashing the potential [M/OL] //TSO The Stationery Office, London: 2012 [2021-07-01]. <https://www.gov.uk/government/publications/open-data-white-paper-unleashing-the-potential>.
- [3] COLAVIZZA G, HRYNASZKIEWICZ I, STADEN I, et al. The citation advantage of linking publications to research data [J]. PLoS ONE, 2020, 15 (4) : e0230416.
- [4] STM Research Data. Share-Link-Cite [EB/OL]. [2021-07-01]. <https://www.stm-researchdata.org/>.
- [5] STUART D, BAYNES G, HRYNASZKIEWICZ I, et al. Whitepaper: Practical challenges for researchers in data sharing [J/OL]. figshare, 2018. <https://doi.org/10.6084/m9.figshare.5975011.v1>.
- [6] WILKINSON M, DUMONTIER M, AALBERSBERG I, et al. The FAIR Guiding Principles for scientific data management and stewardship [J]. Scientific Data, 2016, 3: 160018. <https://doi.org/10.1038/sdata.2016.18>.
- [7] FAIRsharing. Generalist Data Repository Grid [EB/OL]. [2021-07-01]. <https://fairsharing.org/collection/GeneralRepositoryComparison>. DOI: 10.5281/zenodo.3946720.
- [8] Mendeley Data. Dataset Archiving [EB/OL]. [2021-07-01]. <https://data.mendeley.com/>.
- [9] DANS. Dutch National Centre of Expertise and Repository for Research Data [EB/OL]. [2021-07-01]. <https://dans.knaw.nl/en>.
- [10] OpenAIRE [EB/OL]. [2021-07-01]. <https://www.openaire.eu/>.
- [11] DataCite [EB/OL]. [2021-07-01]. <https://datacite.org/>.
- [12] SCHOLIX [EB/OL]. [2021-07-01]. <http://www.scholix.org/>.
- [13] Mendeley Data. File Formats [EB/OL]. [2021-07-01]. <https://data.mendeley.com/file-formats>.
- [14] 翟军, 梁佳佳, 吕梦雪, 等. 欧盟开放科学数据的FAIR原则及启示 [J]. 图书与情报, 2020 (6) : 103-111.
- [15] 邱春艳. 国内外科学数据相关实践中的元数据研究进展 [J]. 情报资料工作, 2021, 42 (5) : 104-112.
- [16] SCHRIML L M, CHUVOCHINA M, DAVIES N, et al. COVID-19 pandemic reveals the peril of ignoring metadata standards [J]. Scientific Data, 2020, 7 (1) . DOI:10.1038/s41597-020-0524-5.
- [17] HRYNASZKIEWICZ I, SIMONS N, HUSSAIN A, et al. Developing a research data policy framework for all journals and publishers [J]. Data Science Journal, 2020, 19 (1) : 5.
- [18] 赵展一, 黄金霞. 开放科学基础设施的信息资源建设模式分析 [J]. 图书馆建设, 2021 (3) : 46-55.
- [19] 黄永文, 张建勇, 黄金霞, 等. 国外开放科学数据研究综述 [J]. 现代图书情报技术, 2013 (5) : 21-27.
- [20] 金贞燕, 阿童木. 科研数据管理服务内容体系构建研究 [J]. 情报理论与实践, 2021, 44 (8) : 42-50.
- [21] 顾立平, 张潇月. 开放科学环境下数据馆员的实践探析 [J]. 图书情报知识, 2020 (2) : 60-74, 112.
- [22] 武彤. 基于数据生命周期的美国研究图书馆科学数据开放共享服务研究 [J]. 图书与情报, 2019 (1) : 135-144.

- [23] 顾立平. 科研模式变革中的数据管理服务: 实现开放获取、开放数据、开放科学的途径 [J]. 中国图书馆学报, 2018, 44 (6): 43-58. Declaration of Data Citation Principles [EB/OL]. [2021-07-01]. <https://doi.org/10.25490/a97f-egykh>.
- [24] 罗鹏程, 崔海媛, 赵静茹. 基于DataCite的科学数据现状特征研究 [J]. 图书情报知识, 2019 (3): 80, 101-112. [26] COUSIJN H, KENALL A, GANLEY E, et al. A data citation roadmap for scientific publishers [J]. Scientific Data, 2018, 5: 180259.
- [25] MARTONE M. Data Citation Synthesis Group: Joint

作者简介

卢焱, 女, 1983年生, 博士, 副研究员, 研究方向: 农业信息资源建设与组织。
王鹭飞, 男, 1973年生, 硕士, 副研究馆员, 研究方向: 农业信息资源建设。
刘洪冰, 女, 1990年生, 硕士, 馆员, 研究方向: 农业信息资源建设。
袁雪, 女, 1984年生, 博士, 馆员, 研究方向: 农业信息资源建设。
陈斯翰, 男, 1996年生, 研究实习员, 研究方向: 信息技术。
刘敏娟, 女, 1982年生, 博士, 副研究员, 通信作者, 研究方向: 农业信息资源建设, E-mail: liuminjuan@caas.cn.

Practice and Enlightenment on Open Scientific Data of Mendeley Data Platform

LU Yao WANG YueFei LIU HongBing YUAN Xue CHEN SiHan LIU MinJuan
(Agriculture Information Institution, CAAS, Beijing 100081, China)

Abstract: Efficient research data integrated discovery and management tools promote open exchange and sharing of research data. The paper investigates into the Elsevier Mendeley Data research data management platform, in the aspects of its data governance processes, integrated open data ecosystem, the functions of data storage, management and sharing, summarizes the advantages in its process design and system development centered around the data standard in its implementation of open data practice, that would provide references for domestic libraries of universities and research institutes and information service institutions to conduct open data resources construction and research data service.

Keywords: Research Data; Open Data; Data Management; Data Sharing; Mendeley Data

(收稿日期: 2021-08-02)