

国内外健康医学科学数据管理平台 对比分析

袁梦雪

(武汉大学信息管理学院, 武汉 430072)

摘要: 基于科学数据管理生命周期理论与科学数据管理能力成熟度模型, 从建设基础(政策、经费、技术、团队)及管理过程(数据管理计划、采集标准与流程、描述与元数据、存储与保存、共享与保护、应用与服务)两个维度对比国内外11个健康医学科学数据管理平台的建设实践。调查发现, 国内健康医学科学数据管理政策不够完善, 缺少保障健康医学数据安全及个人隐私的法规; 平台提供的服务强调健康医学科学数据存储, 而对数据管理与共享的支持不足。基于国外成熟经验, 本文从完善激励机制与数据政策、扩大经费来源、增强技术创新、保护数据安全和促进数据共享、深化服务内容5个方面对国内健康医学科学数据管理平台提出建议。

关键词: 健康医学; 科学数据管理; 数据共享; 健康医学科学数据管理平台

中图分类号: G250

DOI: 10.3772/j.issn.1673-2286.2020.01.002

健康医学科学数据管理平台可以实现健康医学科学数据的采集、组织、存储、管理、分析、共享及发布等^[1]。国内针对科学数据管理平台的研究主要聚焦于两个方向, 一是科学数据管理平台的建设现状和评价标准, 包括国内外科学数据管理平台的建设成就^[2-3]、建设标准规范^[4]以及平台评价体系构建^[5-6]; 二是科学数据管理平台技术研发与应用, 包括平台建设技术^[7]、运行机制^[8]、系统模型以及设计实现^[9]。部分研究人员梳理了国内外健康医学科学数据管理的政策演化与发展^[10-11]; 孙小康等^[12]提出健康医学科学数据管理平台系统设计方案, 如国家人口与健康科学数据共享平台(National Population Health Science Data Center, NCMI)^[13]与国家基因组科学数据中心(the BIG Data Center, BIGD)^[14]的设计; 李翼等^[15]调研了开放医学科学数据平台的构建情况。但当前研究缺乏对健康医学科学数据管理平台的全面考察, 通过对比分析国内外健康医学科学数据管理平台的建设实践可以为国内健康医学科学数据的管理与再利用提供先进经验。基于此, 根据科学数据生命周期理论与科学数据管理能力

成熟度模型(Capability Maturity Model for Scientific Data Management, SDM-CMM), 结合健康医学科学数据的特征, 本文从建设基础与管理过程两个层面对比分析国内外健康医学科学数据管理平台的建设实践, 为国内健康医学科学数据管理平台的建设提供参考与借鉴。

1 研究设计

科学数据管理生命周期涵盖科学数据的产生、收集、描述、存储、共享、应用整个生命过程, 识别了数据管理的核心要素, 主要包括数据管理计划、数据采集、数据处理、数据保存、数据共享与再利用。Crowston等^[16]提出科学数据管理能力成熟度模型, 基于科学数据管理的关键实践, 将科学数据管理划分为特定过程与通用过程。特定过程指与科学数据管理生命周期对应的科学数据管理工作流程; 通用过程指围绕管理或支持科学数据管理过程开展的保证科学数据管理可靠性的准备基础。本文基于科学数据管理生命周期理论与科学

数据管理能力成熟度模型,构建如图1所示的研究框架,从数据政策、经费支持、团队建设和技术基础4个方面对比国内外健康医学科学数据管理平台建设基础的差异;从数据管理计划、数据采集标准与流程、数据描述与元数据、数据共享与保护、数据存储与保存、数据应用与服务6个方面对比国内外健康医学科学数据管理平台管理过程的差异。

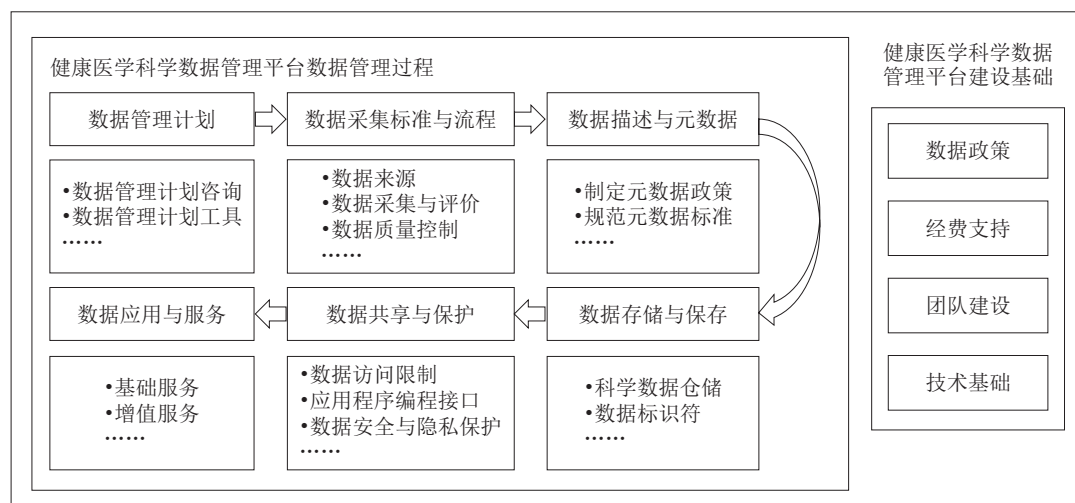


图1 健康医学科学数据管理平台的研究框架

2 平台建设基础对比

2.1 数据政策

健康医学科学数据管理政策的制定者主要有3个主体,即政府、科研资助机构、健康医学科学数据管理平台^[28]。政府政策从宏观上引导和推动科研机构和其他相关部门数据开放管理政策的制定,如美国的《数据获取法令》、英国的《数据保护法案》、澳大利亚的《信息自由法》等,响应政府对数据开放的要求,科研资助机构制定数据管理与共享政策使健康医学科学数据尽可能地以可重复利用的格式开放,促进现有数据的使用^[28]。如美国国立卫生研究院(National Institutes of Health)制定的《数据共享政策和实施指南》(*Data Sharing Policy and Implementation Guidance*)和英国生物技术和生物研究理事会(British Council for Biotechnology and Bioscience Research, BBSRC)制定的《BBSRC数据共享政策》(*BBSRC Data Sharing Policy*)均要求受资助者在申请科研资助时提交数据管理计划和数据共享计划,并且强调数据汇交的必要性,

科学数据仓储目录(the Registry of Research Data Repositories, Re3data)统计了跨学科科学数据仓储标准信息。FAIRsharing重点关注数据、元数据及其在数据仓储中实施的标准。利用Re3data与FAIRsharing,结合期刊出版推荐、访问下载量、文献引用等指标,选择功能完整且建设较成熟的8个国外平台和3个国内平台作为研究对象,详细内容见表1。

要求共享科学数据^[29-30];部分健康医学科学数据管理平台建设主体也基于现有科学数据管理政策,构建了平台自身的数据共享管理政策与使用条例(见表2)。如Brain-CODE制定了信息治理政策与平台使用条款协议,概述了用户如何收集、存储和访问数据,为平台的运营与管理提供了依据^[31-32];TCIA制定了数据使用政策,为数据管理与共享提供了标准规范^[22]。

国内的科学数据管理法规政策已初步形成,但数据管理仍处于条块分割和“各自为政”的状态^[33]。《国家科技资源共享服务平台管理办法》《科学数据管理办法》《中华人民共和国人类遗传资源管理条例》从国家层面上明确科学数据管理的职责、原则、方式和机制,规定强制性汇交,鼓励对科学数据的合理利用^[34-36];但缺乏科研资助机构以及健康医学科学数据管理平台建设主体针对健康医学科学数据的全生命周期管理政策和数据处理规范文件。

2.2 经费支持

国外健康医学科学数据管理平台经费来源广泛,

表1 国内外健康医学科学数据管理平台介绍

平台名称	国家/组织	建设年份	建设主体	建设目标
National Center for Biotechnology Information (NCBI) ^[17]	美国	1988	美国国立生物技术信息中心	利用技术加深对疾病相关的分子及遗传过程的理解
The National Database for Autism Research (NDAR) ^[18]	美国	2007	美国国立卫生研究院	获得数据访问权限, 促进数据共享
The Extensible Neuroimaging Archive Toolkit (XNAT) ^[19]	美国	2007	华盛顿大学医学院神经信息学研究组	规范神经影像数据的常规管理
Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC) ^[20]	美国	2007	加州大学圣地亚哥分校生物系统研究中心	通过发现和比较神经成像研究资源, 促进共享, 减少数据重复, 提高研究质量
Brain-CODE ^[21]	加拿大	2018	安大略省脑科研究院	支持多种脑疾病的多模态数据共享和分析
The Cancer Imaging Archive (TCIA) ^[22]	美国	2004	美国国家癌症研究所 (NCI)	开放信息资源, 利用先进的医学影像学技术, 支持研究、开发和教育活动
Laboratory-European Bioinformatics Institute (EMBL-EBI) ^[23]	欧盟	1994	欧洲分子生物学实验室	通过提供生物数据、工具和知识, 向行业推广尖端技术, 促进生命科学研究的转化
International Network for the Demographic Evaluation of Populations and Their Health (INDEPTH) ^[24]	国际人口与健康人口统计评估网络	1998	国际人口与健康人口统计评估网络	为卫生和人口监测系统及时提供纵向证据, 改善人口健康发展政策与实践
NCMI ^[25]	中国	2010	中国医学科学院医学信息研究所	为政府决策、人口健康、医疗卫生、科技创新、产业发展提供权威、开放、便捷的数据共享和信息服务
国家基因库生命大数据平台 (CNGB) ^[26]	中国	2018	华大集团大数据中心大数据应用团队	促进生命多组学大数据的互联互通、开放共享及有效利用, 推动生命科学和生物产业的快速发展
国家基因组科学数据中心 (BIGD) ^[27]	中国	2019	中国科学院北京基因组研究所	提供全面、高质量的资源, 促进数据开放共享, 保障数据安全可控, 支撑国家科技创新和社会经济发展

包括政府科研基金、基金会赞助、私营企业投资、高校资助等。此外, 部分平台通过提供咨询服务, 以及为其他机构或企业提供元数据等方式获得外部资金支持。如EMBL-EBI的行业计划会与全球商业公司的主要利益相关者和意见领袖定期保持联系和互动, 并通过提供必要的数据库基础架构和服务来支持中小型企业的发展^[23]。国内健康医学科学管理平台的建设资金主要依托于政府机构, 来源单一。如NCMI的建设经费主要来自科技部科学数据共享工程重大项目资金投入^[25], BIGD资金来源于国家自然科学基金和中国科学院^[27]。

2.3 团队建设

部分国外健康医学科学数据管理平台设计了覆盖数据管理生命周期、职责清晰、职权明确的多层次管理体系, 包括技术开发团队、信息咨询团队、政策信息团

队、数据访问管理团队、数据隐私保护团队和用户服务团队。如表2所示, NCMI、CNGB和BIGD都构建了多层次的管理体系。国内平台的管理团队构建已初见规模, 具体分工等细节还需进一步完善。

2.4 技术基础

健康医学科学数据管理平台的技术开发模式主要分为自主研发(包括合作研发)、开源二次开发以及委托第三方开发3类。国外平台以自主研发为主, 如NCBI自主研发了平台系统及生物数据搜索软件Entrez; 由华盛顿大学医学院神经信息学研究组开发的XNAT平台, 也被嵌入NITRC与Brain-CODE管理神经影像学数据。国内平台也以自主研发为主, 部分平台为开源二次开发, 如北京大学开放研究数据平台是基于Dataverse二次开发的。但健康医学科学数据管理的专业技术人

表2 国内外健康医学科学数据管理平台建设基础对比

平台名称	数据政策				经费来源					开发模式			多层次的管理体系
	数据管理政策	数据共享政策	数据使用政策	数据安全与隐私策略	政府机构	科研机构(研究所)	高校	企业	非营利性组织	自主研发	开源二次开发	第三方委托	
NCBI	√	√	√	√	√					√			√
NDAR	√	√	√	√	√	√				√			√
XNAT	√			√	√	√	√			√			
NITRC	√		√	√	√	√	√			√			√
Brain-CODE	√	√	√	√	√	√				√			√
TCIA		√	√	√	√	√	√			√			√
EMBL-EBI	√	√	√	√	√	√		√	√	√			√
INDEPTH	√	√	√	√	√	√	√	√	√	√			√
NCMI	√	√			√					√			√
CNGB			√	√	√		√			√			√
BIGD			√	√	√					√			√

注：“√”表示平台具备对应的建设基础

才较为匮乏，平台构建的整体技术水平存在局限性。

3 平台数据管理过程对比

3.1 数据管理计划

数据管理计划描述了项目期间以及项目结束后的数据处理方式，通常覆盖数据生命周期的全部或部分，即从数据发现、收集和组织，质量保证/质量控制、文档说明与数据使用，到数据存储与共享管理^[37]。数据管理计划是健康医学科学数据管理的重要环节，NCBI、EMBL-EBI、NDAR、INDEPTH以及TCIA都为研究人员制订健康医学科学数据管理计划提供服务。而国内平台均缺少这一环节，未形成覆盖全生命周期的健康医学科学数据管理。

3.2 数据采集标准与流程

3.2.1 数据来源

国外平台的健康医学科学数据来源广泛。包括NCBI、NDAR、NITRC、TCIA、EMBL-EBI在内的5个平台均是被多家期刊出版商推荐的数据库，其数据来源不受任何地域和机构限制。此外，数据联合有助于健康医学科学数据的集成，拓宽了健康医学科学数据管理平台的数据来源。如Brain-CODE与NDAR积极开

展联合研究，增加数据量并提供更丰富的分析资源。国内平台的数据来源相对受限。仅有BIGD的组学原始数据归档库（Genome Sequence Archive, GAS）被国际知名出版商Elsevier收录为指定基因数据归档库。受采集成本、科研绩效、数据安全、共享政策等多方限制，国内平台数据来源集中在合作的高校与科研机构。以NCMI为例，其数据主要来自10个合作机构，外源数据占比小^[38]，且平台数据在及时更新与补充上难以得到保证。

3.2.2 数据采集与评价

数据采集与评价的核心任务是对健康医学科学数据价值进行判断与筛选，包括新数据集收录前的鉴定和评价，以及对现有数据集的筛选与再评价^[39]，有助于提高健康医学科学数据管理效率。国外平台制定了不同的采集评价流程，XNAT、TCIA和EMBL-EBI开发有自动数据采集工具，具体见表3。部分平台有多个科学数据库，如NCBI和EMBL-EBI，各个数据库有不同的数据采集与评价标准。国内平台的采集流程中涉及较多手工操作，如由用户填写数据标准（元数据）项。

3.2.3 数据质量控制

国内外11个健康医学科学数据管理平台采集流程中均包含质量控制。国外8个平台均制定了全面的质

表3 国内外健康医学科学数据管理平台采集流程

平台名称	采集流程
NCBI	多种采集工具，如BankIt、Tb12asn、The Sequence Read Archive (SRA) 等。以SRA为例： ①新建提交，注册项目或样本；②提交SRA元数据，上传序列数据文件
NDAR	①获取参与者的同意，以及参与者的个人健康数据（生成全局唯一标识符的资料）；②基于成本模型，估算数据共享成本； ③完成数据提交协议；④根据NDAR数据字典，定义需提交的数据结构；⑤完成数据验证，并提交数据
XNAT	①从XNAT网站下载安装Upload助手小程序，提交数据；②XNAT Upload读取DICOM元数据， 构造并验证会话元数据，并继续去标识与匿名化操作
NITRC	①阅读NITRC用户指南，了解数据必须满足的规范要求和可用的管理工具；②在NITRC注册页面上填写数据的 必要信息，选择源代码管理系统；③NITRC站点管理员批准后，设置数据的管理方案
Brain-CODE	①各方签署参与协议并加送至OBI后，平台管理人员将为协议列出的参与者创建账号；②用户获得数据采集工具， 如REDCap、LabKey和SPReD的使用权限，对采集的数据进行去标识与加密操作；③依据Brian-CODE 通用数据元素（Common Data Elements, CDE），采集并提交数据
TCIA	①提交数据申请，由TCIA咨询小组进行审核；②通过审核后，下载安装临床试验处理器（Clinical Trials Processor, CTP） 和POSDA（Perl Open Source Digital Imaging and Communications in Medicine Archive）工具，提交数据；③CTP与 POSDA检测数据，完成去标识操作；④TCIA质量控制和策划人员与用户合作确保完全识别并接收数据，发布最终数据集
EMBL-EBI	多种采集方式，包括交互方式、程序方式、Webin方式等。以Webin为例： ①注册提交用户，从网站上下载安装Webin软件；②预注册研究与样本，准备要上传的文件，包括清单文件格式， 文件类型与信息文件；③验证并提交文件
INDEPTH	-
NCMI	①用户新建元数据，上传数据；②申请数据证书；③审核通过后平台发布最终数据集
CNGB	①阅读CNGB的数据提交手册；②进入提交入口，按照要求填写健康医学科学数据信息； ③依据提交模板，提交数据并进行校验
BIGD	①阅读BIGD的使用指南，了解数据标准；②登录BIG Sub平台，根据网页信息提示提交数据

量检查与控制流程，保证数据的准确性与真实性。如NDAR的双层质量控制，第一级在数据提交前，包括自动化工具的项目级验证与人工数据准确性验证；第二级在数据提交后，由数据委员会审查资料的完整性、准确性和质量，生成质量审查报告^[18]。而国内健康医学科学数据管理平台的质量控制与管理环节中，由平台管理人员对提交的数据进行质量审核，平台未提供自动化质量检测工具，也未提供质量检查报告帮助科研人员控制健康医学科学数据质量。

3.3 数据描述与元数据

INDEPTH采用的是常用于社会科学信息组织的DDI元数据标准，其他7个国外平台的元数据标准均是平台以现有的元数据标准为基础修改定制的。例如，Brain-CODE开发的通用数据格式（Common Data Elements, CDEs），提供了标准的数据定义与格式和详细的文档说明^[21]，以便研究人员在各类研究中收集数

据的一致性。国内NCMI制定了《人口健康科学数据共享数据集元数据标准》^[13]，BIGD与CNGB整合了国际组学、健康、医学领域的数据结构和标准，构建了具有广泛兼容性、标准化的数据标准和结构，方便与外部数据平台链接。

3.4 数据存储与保存

除了NCMI以外的10个平台都在Re3data平台上为对应的科学仓储进行注册，便利科研人员使用。国外平台中，NDAR、TCIA和INDEPTH为数据分配了数字对象标识符（Digital Object Identifier, DOI），EMBL-EBI为数据分配了科研工作者标识符（Open Research and Contributor ID, ORCID），使得数据更易被找到、复用和长期保存。国内CNGB支持DOI，NCMI为数据分配了系统生成的唯一标识符，BIGD的DOI功能正在开发中。

3.5 数据共享与保护

3.5.1 数据共享

国内外11个平台基本都采取了数据分级获取、数据拥有者自定共享模式（开放/受控）、分层开放访问等方式支持数据共享，具体见表4。访问与上传限制均以注

册式限制为主，即注册用户才可上传与访问数据；应用程序编程接口（Application Programming Interface，API）为用户提供高吞吐量的数据访问。国外平台会提供多类API接口，方便有不同技术习惯的科研团队下载平台数据。国内平台支持的API接口类型单一。国内外平台的共享程度差别较大，国外平台的每日下载量与访问量远超国内平台。

表4 国内外健康医学科学数据管理平台数据共享情况

平台名称	支持语言	数据访问类型		数据上传类型		API
		开放型	限制型	开放型	限制型	
NCBI	英文	√			√	FTP
NDAR	英文		√		√	-
XNAT	英文	√			√	FTP
NITRC	英文	√	√		√	SOAP
Brain-CODE	英文		√		√	√
TCIA	英文	√	√		√	REST
EMBL-EBI	英文	√	√	√		FTP; REST; SOAP
INDEPTH	英文		√		√	-
NCMI	中文	√			√	-
CNGB	英文和中文	√			√	FTP
BIGD	英文和中文	√			√	FTP

注：“√”表示平台支持对应的功能类型

3.5.2 数据安全与隐私保护

健康医学科学数据涉及大量个人敏感信息、隐私数据、科研机密等内容^[40]，国外平台从法律政策和平台管理两个层面采取措施加强对敏感信息的保护。法律政策层面，国外已出台系列相关政策，如美国的《健康保险流通与责任法案》（*Health Insurance Portability and Accountability Act, HIPAA*）以及欧盟的《通用数据保护条例》（*General Data Protection Regulation, GDPR*）规范了个人隐私信息的使用、披露与保护。国外8个平台均制定了专门的数据安全与隐私保护政策，或在数据管理政策中说明用户健康医学科学数据与个人数据的使用规范，涵盖用户知情同意、个人信息采集与利用、持久性Cookie相关技术、个人信息储存与披露、平台安全防护等方面。平台管理层面，国外8个平台均采用数据去标识技术（清除个人信息）、数据加密技术（对敏感信息模糊处理）以及数据组织分层+访问控制（限制科研人员获取数据的权限）等方式保障数据安全。如Brain-CODE提供了持续监控软件，识别未授权的访问，

并定期进行隐私影响评估和威胁风险评估。

国内的《信息安全技术个人信息安全规范》基于个人信息的全生命周期，提出了保护个人信息应遵循的原则与安全要求，但缺少对数据安全与隐私保护的识别、分类和监管等内容。国内的3个平台中，NCMI的管理文件缺失了数据安全与隐私保护的内容，BIGD在数据使用政策中提到了隐私政策，只有CNGB制定了专门的隐私和安全政策。

3.6 数据应用与服务

根据用户需求，可将平台服务划分为基础服务与增值服务。基础服务是健康医学科学数据管理平台为满足用户基本需求所提供的基础性服务，包括用户指南/手册、数据采集上传、数据组织存储、数据访问检索、数据共享下载。而增值服务是为了满足用户更深层次的需求，其能有效提高平台的易用性与用户满意度。国外平台提供了形式多样的服务，包括检索优化、数据分析、整合集成、知识发现、公告发布、计算资源、可视

化、交流互动和教学培训等。表5显示，Brian-CODE与TCIA提供数据多维可视化服务，以最直观的方式展示数据的分布与特征。NITRC支持资源工具的比较，便于用户选择。作为数据提供平台，NCMI以提供数据存储与下载服务为主，为用户提供咨询服务。CNGB与BIGD正积极利用现有数据资源，为健康医学科学数据的分析、挖掘与再利用提供方便。

表5 国内外健康医学科学数据管理平台服务情况

平台名称	基础服务					增值服务								
	用户指南/手册	数据采集上传	数据组织存储	数据访问检索	数据共享下载	检索优化	数据分析	整合集成	公告发布	教学培训	计算资源	可视化	交流互动	知识发现
NCBI	√	√	√	√	√	√	√	√	√	√		√	√	√
NDAR	√	√	√	√	√	√	√	√	√	√	√			
XNAT	√	√	√	√	√	√	√	√	√	√	√		√	
NITRC	√	√	√	√	√	√	√	√	√		√		√	√
Brain-CODE	√	√	√	√	√	√	√	√		√	√	√		
TCIA	√	√	√	√	√	√	√	√	√			√		
EMBL-EBI	√	√	√	√	√	√	√	√	√	√	√			√
INDEPTH	√	√	√	√	√	√	√	√	√	√				
NCMI	√	√	√	√	√				√			√		
CNGB	√	√	√	√	√	√	√	√	√					
BIGD		√	√	√	√	√	√	√	√	√				

注：“√”表示平台提供对应的服务

4 启示与建议

本文在对国内外健康医学科学数据平台进行深入对比分析的基础上，从健康医学科学数据的数据政策与激励机制、经费与数据来源、技术创新与人才培养、数据安全与共享、平台服务的角度，总结了国外健康医学科学数据管理平台的建设经验。结合国内健康医学科学数据管理平台建设的不足，对国内提升平台建设水平提出以下建议。

4.1 完善数据政策, 建立激励机制

一方面，我国需进一步完善健康医学科学数据管理系列法律法规，在《科学数据管理方法》宣示性条文的基础上，制定更多授权性和禁止性条文来规范健康医学科学数据的知识产权、数据存储、安全保护、数据出版、管理监督等相关实践，形成一套完整、健全、可执行的健康医学科学数据管理政策体系；另一方面，基于《科学数据管理办法》，健康医学相关的科研资助机构和科研机构，如健康医学研究所、地方医院、高校图书馆、期刊出版社等须根据健康医学数据需求，制

定健康医学科学数据管理与共享办法，提出数据汇交的明确要求，规定健康医学科学数据如何进行采集、管理、保存、共享与利用，加强健康医学科学数据整合，规范健康医学科学数据管理。此外，政府、科研资助机构、健康医学研究机构 and 高校应当建立激励机制，以资金、数据优先使用权、科研业绩或影响力等方式作为奖励，鼓励项目人员主动共享研究中产生的健康医学科学数据。

4.2 强化经费支持, 拓宽数据来源

国内健康医学科学数据管理平台建设的主要经费来源于基金和项目经费，然而单一的经费来源难以支撑平台的可持续发展，还需要社会其他力量的支持。借鉴国外平台的有偿服务经验，有偿提供计算资源、数据分析报告、行业咨询、决策支持等服务，吸引来自社会组织机构、企业或个人的资助。鼓励地方政府、高校、科研院所与企业对健康医学科学数据管理平台建设的支持，促进各方协同合作，共同收集、管理和共享健康医学科学数据，加大外源数据采集力度，拓展健康医学科学数据来源。

4.3 增强技术创新, 加强人才培养

国内平台构建较多采用自主开发方式, 其开发的软件少有开源, 而开源的构建软件可有效缩短健康医学科学数据管理平台开发周期, 节约开发成本。建议以中国科学院为首的国内平台开发团队, 开发并开源类似DSpace、DataVerse、CKAN的平台构建软件, 降低平台的建设难度, 推动平台服务的普及; 开发面向科研全过程、覆盖数据全生命周期的通用工具, 如自动健康医学科学数据采集工具能提高健康医学科学数据管理规范性与自动水平。此外, 还应加强健康医学科学数据管理人才的培养, 扩大人才储备。

4.4 保证数据安全, 促进数据共享

国内健康医学科学数据安全性与隐私保护的法规建设还不够成熟。科研资助机构、高校以及健康医学科研机构需完善健康医学科学数据的隐私保护与安全保障机制, 确立一套完整的健康医学科学数据的安全评估标准, 并制定对应的违规处理办法, 遏制威胁数据安全的行为。健康医学科学数据管理平台应在确保遵守道德原则和安全法规的方式采集存储数据的前提下, 促进数据的下载共享。数据质量良莠不齐是影响数据共享的重要因素, 平台需加强质量控制, 提高数据质量, 加大宣传推广, 提高知名度, 推动数据共享。

4.5 扩宽服务体系, 深化服务内容

国内健康医学科学数据管理平台建设时间较晚, 增值服务较为匮乏。拓展平台的服务, 需要增设数据管理计划, 借鉴英国数字监管中心(Digital Curation Centre)的数据管理计划模板经验, 设计一个通用的模板来规范健康医学科学数据管理计划的内容, 利用已开发的DMPonline^[41], 增加数据管理计划制定咨询服务; 增加平台的增值服务, 完善数据可视化与数据分析服务, 增强数据分析功能, 提供更丰富的在线分析服务; 考虑在平台中添加论坛讨论功能, 促进用户以及平台管理人员与数据使用者之间的沟通交流, 提高用户积极性, 扩大平台影响力; 加大平台管理人员培训力度, 扩大咨询人员数量, 提供更多个性化的健康医学科学数据管理服务。

参考文献

- [1] 关于促进和规范健康医疗大数据应用发展的指导意见[EB/OL]. [2019-12-16]. http://www.gov.cn/zhengce/content/2016-06/24/content_5085091.htm.
- [2] 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): 97-107.
- [3] 张莎莎, 黄国彬, 耿寒. 基于re3data的英国科学数据发布平台研究[J]. 数字图书馆论坛, 2017(6): 16-24.
- [4] 赵华, 王健. 国内外科学数据元数据标准及内容分析[J]. 情报探索, 2015(2): 21-24.
- [5] 司莉, 李月婷, 邢文明, 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.
- [6] 刘桂锋, 张裕, 刘琼. 科研数据开放平台评价指标体系构建及案例研究[J]. 图书情报知识, 2019(1): 21-31.
- [7] 张计龙, 殷沈琴, 张用, 等. 社会科学数据的共享与服务——以复旦大学社会科学数据共享平台为例[J]. 大学图书馆学报, 2015, 33(1): 74-79.
- [8] 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.
- [9] 高芹, 钟晓莉. 高校图书馆科学数据监护平台构建研究[J]. 图书馆学刊, 2015(8): 113-116.
- [10] 唐源, 吴丹. 国外医学科学数据共享政策调查及对我国的启示[J]. 图书情报工作, 2015, 59(18): 6-13.
- [11] 杨林, 李姣, 杨啸林, 等. 我国卫生健康领域科学数据全链条管理现状调查与分析[J]. 图书情报工作, 2017, 61(22): 117-126.
- [12] 孙小康, 吴思竹, 张泽, 等. 医学科学数据管理与共享平台原型系统建设[J]. 医学信息学杂志, 2018, 39(8): 19-24.
- [13] 姜吉栋, 赵辉, 刘润达. 科学数据共享平台网站中的信息组织——以国家人口与健康科学数据共享平台为例[J]. 信息资源管理学报, 2012(4): 52-56.
- [14] 张源笙, 夏琳, 桑健, 等. 生命与健康大数据中心资源[J]. 遗传, 2018, 40(11): 1041-1045.
- [15] 李翼, 吴丹. 开放医学科学数据平台调查研究[J]. 图书情报工作, 2015, 59(18): 24-29.
- [16] CROWSTON K, QIN J. A capability maturity model for scientific data management: Evidence from the literature[J]. American Society for Information Science and Technology, 2011, 48(1): 1-9.
- [17] National Center for Biotechnology Information[EB/OL]. [2019-12-16]. <https://www.ncbi.nlm.nih.gov/>.
- [18] The National Database for Autism Research[EB/OL].

- [2019-12-16]. <https://ndar.nih.gov/>.
- [19] The Extensible Neuroimaging Archive Toolkit [EB/OL]. [2019-12-16]. <https://www.xnat.org/>.
- [20] Neuroimaging Informatics Tools and Resources Clearinghouse [EB/OL]. [2019-12-16]. <https://www.nitrc.org/>.
- [21] Brain-CODE [EB/OL]. [2019-12-16]. <https://www.braincode.ca/>.
- [22] The Cancer Imaging Archive [EB/OL]. [2019-12-16]. <https://www.cancerimagingarchive.net/>.
- [23] European Molecular Biology Laboratory-European Bioinformatics Institute [EB/OL]. [2019-12-16]. <https://www.ebi.ac.uk/>.
- [24] International Network for the Demographic Evaluation of Populations and Their Health Data Repository [EB/OL]. [2019-12-16]. <http://www.indepth-network.org/>.
- [25] 国家人口健康科学数据中心 (NCMI) [EB/OL]. [2019-12-16]. <http://www.ncmi.cn/index.html>.
- [26] 国家基因组生命大数据平台 (CNGB) [EB/OL]. [2019-12-16]. <https://db.cngb.org/>.
- [27] 国家基因组科学数据中心 (BIGD) [EB/OL]. [2019-12-16]. <https://bigd.big.ac.cn>.
- [28] 刘莉, 刘文云, 刘建, 等. 英国科研数据管理与共享政策研究 [J]. 情报资料工作, 2019, 40 (5): 46-53.
- [29] NIH. NIH Data Sharing Policy and Implementation Guidance [EB/OL]. [2019-12-16]. https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm.
- [30] BBSRC. BBSRC Data Sharing Policy [EB/OL]. [2019-12-16]. <https://bbsrc.ukri.org/documents/data-sharing-policy-pdf/>.
- [31] BRAIN-CODE. Brain-CODE Governance Policies [EB/OL]. [2019-12-16]. https://www.braincode.ca/sites/default/files/about/OBI-Governance_v2.1_09JAN2019.pdf.
- [32] BRAIN-CODE. Platform Terms of Use Agreement [EB/OL]. [2019-12-16]. <https://www.braincode.ca/content/platform-terms-use-agreement>.
- [33] 周玉琴, 邢文明. 我国科研数据管理与共享政策体系研究 [J]. 中华医学图书情报杂志, 2018, 27 (8): 1-7.
- [34] 中华人民共和国人类遗传资源管理条例 [EB/OL]. [2019-12-16]. http://www.gov.cn/zhengce/content/2019-06/10/content_5398829.htm.
- [35] 科技部财政部关于印发《国家科技资源共享服务平台管理办法》的通知 [EB/OL]. [2019-12-16]. http://www.most.gov.cn/mostinfo/xinxifenlei/fgzc/gfxwj/gfxwj2018/201802/t20180224_138207.htm.
- [36] 国务院办公厅关于印发科学数据管理办法的通知 [EB/OL]. [2019-12-16]. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.
- [37] MICHENER W K. Ten Simple Rules for Creating a Good Data Management Plan [J]. PLOS Computational Biology, 2015, 11 (10): e1004525.
- [38] NCMI. NCMI资源统计 [EB/OL]. [2019-12-16]. <http://www.ncmi.cn/statistics/resourceStatistics.html?active=resourceStatistics>.
- [39] 崔旭, 韩子鹤, 王铮, 等. 科学数据管理中的数据选择与评价关键要素研究 [J]. 情报理论与实践, 2018, 41 (3): 78-82.
- [40] 蔡韶莹. 医学院校图书馆面向附属医院开展科学数据管理服务研究 [J]. 大学图书情报学刊, 2019, 37 (4): 18-21.
- [41] DONNELLY M, JONES S, JOHN W. DMP Online: the digital curation centre's web-based tool for creating, maintaining and exporting data management plans [J]. The International Journal of Digital Curation, 2010, 5 (1): 187-193.

作者简介

袁梦雪, 女, 1995年生, 硕士研究生, 研究方向: 健康医学信息、科学数据管理, E-mail: 18807162095@163.com。

Comparative Analysis of Health Medical Scientific Data Management Platforms from Domestic and Abroad

YUAN MengXue

(School of Information Management Wuhan University, Wuhan 430072, China)

Abstract: Based on the scientific data management life cycle theory and the scientific data management capability maturity model, we compare the construction practices of 11 health medical scientific data management platforms from two dimensions of the construction foundation (policy, funding, technology, organization) and management process (data management plan, collection and evaluation, description and metadata, storage and preservation, sharing and protection, applications and services). The survey found that domestic health medical scientific data management policies are inadequate, lacking regulations to protect the health medical data security and personal privacy, and that the services provided by the platform emphasize the storage of health medical scientific data, lacking attention to data management and sharing. According to abroad experience, this article proposes domestic health medical scientific data management platforms from five aspects: improving incentives and data policies, expanding funding sources, enhancing technological innovation, protecting data security and promoting data sharing, and deepening service content.

Keywords: Health Medicine; Scientific Data Management; Data Sharing; Health Medical Scientific Data Management Platforms

(收稿日期: 2019-12-26)