

图书馆数据管理服务设计

——基于中国科学院大学研究生的调查

刘金亚 顾立平

【摘要】图书馆数据管理服务经过长时间的发展与实践,其服务内容和形式在不断地充实。采用半结构化访谈的方法,对中国科学院大学的研究生进行了调查。基于访谈结果的扎根理论分析及文献综述,设计了图书馆数据管理服务的框架和要点。图书馆未来的数据管理服务要以数据权益管理为核心与前提,关注数据管理技能培养的服务效果,帮助克服数据管理流程中的伦理问题,实现数据管理价值的最大化。

【关键词】数据管理服务 科研数据共享 影响因素 扎根理论

Abstract: The library data management services have been developed and practiced for a long time, and its service content and service form have been constantly enriched. This research investigated the postgraduate students in University of Chinese Academy of Sciences by semi-structured interview. Based on the grounded theory analysis of the interview results and literature review, the framework and key points of library data management services are designed. It is concluded that the future data management services of library should take the management of data rights as the core and premise, pay more attention to the service effect of data management skill training, help to solve the ethical problems in the process of data management, and maximize the value of data management.

Key words: data management service research data sharing influencing factors grounded theory

DOI:10.15941/j.cnki.issn1001-0424.2020.18.007

0 引言

各种大型试验设备的改进,国家基础设施的不断完善,使科研数据逐渐成为辅助科研人员完成重大突破、发现科研目标越来越重要的手段。继实验科学、理论科学、计算科学后,科学已进入了数据驱动的第四科研范式^[1]。科研人员通过地面观测、网络抓取、实时监测等手段形成数据,而这些数据的汇集、管理及再利用等工作则由于科研人员的数据需求及其共享行为的不同存在较大差异^[2]。研究生作为未来科学事业的重要参与者,将会在未来科技发展中占据重要作用,是科研数据开放共享重要参与者,调查其科研数据管理服务需求的影响因素,可反映未来科研人员的数据管理服务需求,可帮助图书馆设计科研数据管理新服务,帮助实现数据管理价值的最大化。

1 研究背景

数据管理最早出现在美国国家科学基金会(National Science Foundation,简称NSF)提出的研究报告《通过信息基础设施促进科学和工程的革命(Revolutionizing Science and Engineering through Cyberinfrastructure)》中^[3]。随着数据管理的不断发展,世界各国相继发布了数据管理政策。其中包括美国科学基金会的数据管理与共享政策^[4]、英国研究与创新理事会的数据政策^[5]、中国的科研数据管理办法^[6]等。

为了保障科学数据管理服务工作的顺畅,各个组织搭建了基础设施。主要包括:欧盟的欧洲开放科学云^[7]和OpenAIRE^[8],美国的康纳尔大学的arXiv预印本平台^[9]与哈佛大学的Dataverse数据存储平台^[10],澳大利亚的ANDS数据服务平台^[11],日本国立情报研究所的SINET网络基础设施^[12],中国科学院的科技云^[13]、CASEarth^[14]和DBAR^[15]。

此外,数据管理服务实践活动也非常丰富。美国耶鲁大学、康纳尔大学等成立专门的数据管理服务中心,为用户提供科学数据管理计划、共享、存储、检索、导航、分析、咨询、元数据、管理和培训、相关政策等有关的服务内容^[16]。Springer Nature^[17]围绕全球科研人员的需求展开了与数据发现有关的一系列的服务工作。

我国科研管理数据服务大概经过了初探、实践、创新三个阶段（表1）。

表1 我国数据管理服务的发展阶段

阶段	主要研究内容	描述
初探阶段	数据管理服务的含义 ^[18] 、服务架构 ^[19-20] 、机遇与挑战 ^[21] 、国际现状调研 ^[22] 等	这一阶段的研究内容主要集中在数据管理服务是什么、服务架构有什么样的内容
实践阶段	权益管理 ^[23-25] 、数据政策 ^[26] 、基础设施 ^[27-28] 、数据馆员 ^[29-31] 、人才培养与教育 ^[32-34] 、标准与规范 ^[35-36] 、管理能力 ^[37-38] 、管理手段 ^[39-40] 、风险管理 ^[41] 、合作与联盟 ^[42-43] 等	这一阶段的研究内容主要是对数据管理服务过程中的处理对象、应具备的服务能力、手段、技术、政策、如何对数据管理进行安全管控、人才培养等的探讨
创新阶段	服务评估 ^[44-45] 、科研数据生命周期 ^[46-47] 、创新服务 ^[48-49] 等	这一阶段的研究内容体现在现有研究内容和实践成果，结合目前的数据管理服务现状，寻找可能的解决数据管理服务困境的方法

数据管理服务是图书馆业务服务的重要内容，但学术界对可预见未来科研人员的科研数据管理服务需求调研，提前预知未来图书馆服务中所需要进行的与数据管理服务业务部署有关的用户调研还较为缺乏，对数据管理服务对象的调研也还有很多值得探索的问题和可拓展的空间。

2 研究过程

2.1 研究设计

本研究选取中国科学院大学的研究生作为调查对象，采用半结构化访谈方法，调查这一群体的科研数据管理服务需求影响因素。基于调查结果分析和文献综述，设计未来图书馆数据管理服务时的层级和要点，以更好地实现数据管理的价值。

2.1.1 访谈对象

作为未来科研事业的主要参与者，在正式进入科学研究领域前，提前了解并培养研究生的数据管理素养，有利于图书馆评估目前数据管理服务效果，提前部署数据管理服务。为达到该目的，中国科学院文献情报中心开设《数据科学 R 与 Python 实践》课程。而选修此课程的学生，所接触到的数据来自于其所在课题组的一手资料，对数据管理方法需求迫切。本文随机选取 78 名选修该课程的研究生，作为访谈对象。并于课程中期，在学生对数据管理内容具备相关知识的前提下，开展半结构化访谈。其中，调查对象主要来自于中国科学院各个研究所，包括北京基因组研究所（1）、成都山地灾害与环境研究所（1）、成都生物研究所（1）、城市环境研究所（4）、大气物理研究所（1）、地理科学与资源研究所（5）、地质与地球物理研究所（1）、动物研究所（1）、高能物理研究所（1）、国家空间科学中心（1）、国家天文台（1）、云南天文台（2）、过程工程研究所（3）、海洋研究所（1）、计算机网络信息中心（2）、计算机与控制学院（3）、计算技术研究所（2）、昆明动物研究所（1）、力学研究所（1）、南京天文光学技术研究所（1）、南京土壤研究所（1）、沈阳计算技术研究所（8）、沈阳应用生态研究所（2）、生态环境研究中心（2）、声学研究所（1）、数学与系统科学研究院（1）、文献情报中心（4）、武汉植物园（3）、西北生态环境资源研究院（1）、西藏大学（1）、心理研究所（2）、新疆生态与地理研究所（4）、信息工程研究所（1）、亚热带农业生态研究所（1）、烟台海岸带研究所（1）、遥感与数字地球研究所（1）、遗传与发育生物学研究所（2）、植物研究所（4）、智能技术卓越创新中心（1）、中国测绘科学研究院（1）、自动化研究所（2）。

2.1.2 研究方法

（1）扎根理论。扎根理论通过识别受测者的关注重点，通过对访谈结果的步骤化描述，最终可形成理论模型。通过对内容的分析结果，可以得出更加全面和科学的理论模型，更全面地筛选与研究主题相关的概念，构建更加真实、可持续的数据管理服务需求影响模型，更精准地为调查对象提供数据管理服务。

（2）半结构化访谈。半结构化访谈是按照粗线条式的访谈提纲，对调查对象进行面对面的调查方法。调查过程更加准确，调查结果更加细致和全面。本文采取一对多的访谈方式，给每个访谈对象发放了访谈提纲，并就每个问题的回答形式和方向，给予了详细的解释，访谈时间 150 分钟。

调查内容侧重于描述研究生对科研数据管理服务的需求,按照统一的访谈提纲对课程的77名研究生进行半结构化的访谈,具体访谈提纲见表2。数据处理按照扎根理论的处理方法,总结理论模型^[2]。其中,在访谈记录中随机选取2/3的访谈记录进行本文编码,剩下的1/3作为理论饱和度检验。

表2 半结构化访谈提纲

序号	内 容
1	您从事的研究以及所需要的数据来源
2	您的数据采集途径
3	您的数据处理过程
4	您的数据分析方法
5	您是否考虑发表研究数据,以及理由

2.2 编码过程

2.2.1 开放式编码

本文采用逐字编码和录入,进而初始概念化。为了保证资料内容的真实性,选取编码尽量使用原句。其中,在选取标签时,剔除了一些简略、模糊的回答,最后提取原句中初始概念,使用 a_n 表示。随后进行范畴化,并使用 A_n 表示,初始概念提取实例见表3。

表3 初始概念提取实例

初始概念	原始资料语句
a_0 数据库网站	生物信息学研究,课题为基因组的组装和注释,数据来源于美国NCBI数据库网站
a_1 直接采集 a_2 网络下载	基因组数据可以通过测序得到,也可以直接从NCBI下载
a_3 数据标注 a_4 数据清洗	在Linux环境下,使用生物信息学软件工具对基因组进行拼接、组装、注释
a_5 分析软件 a_6 数据分析	基于隐马尔可夫模型的原理,利用R或Python语言编程来分析基因组数据
a_7 积极发表 a_8 产出需求	研究数据本身是全球共享,研究结果要发表,因为实验室需要研究产出

在进行初始概念提取后,对初始概念进行了范畴化,最终得到了79条初始概念,14个范畴(表4)。

表4 开放式编码

范畴化	概念化
A_1 公开数据	a_0 数据库/ a_1 机构网站/ a_{17} 网页下载/ a_{26} 数据开放网站/ a_{40} 网站下载/ a_{45} 统计公报/ a_{46} 统计年鉴/ a_{63} 公开的网络词库/ a_{65} 论文数据
A_2 非公开数据	a_{15} 临床医疗数据/ a_{16} 公司获取/ a_{29} 观测数据/ a_{32} 公司年报数据/ a_{48} 非公开保密数据/ a_{54} 自费花钱/ a_{66} 实验数据/ a_{74} 索取数据
A_3 数据处理软件	a_3 python/ a_4 R/ a_5 excel/ a_6 Citespace/ a_7 vosviewer/ a_8 spss/ a_9 stata/ a_{62} 爬虫工具
A_4 数据处理方法	a_2 手工清洗/ a_{20} 主成分分析/ a_{21} 空间分析/ a_{22} 相关分析/ a_{23} 回归分析
A_5 基本素养	a_{24} 数据分组/ a_{25} 数据建模/ a_{59} 数据科学课程/ a_{78} 撰写数据管理报告
A_6 数据管理策略	a_{10} 科学数据管理政策/ a_{11} 数据共享管理办法
A_7 数据管理过程	a_{19} 缺失值处理/ a_{27} 数据导入/ a_{33} 数据可视化/ a_{47} 问卷调查
A_8 数据处理能力	a_{31} 未接触数据处理/ a_{41} 数据处理错误/ a_{44} 人工搜集/ a_{68} 未接触大量数据处理
A_9 共享行为习惯	a_{12} 考虑全部发表/ a_{30} 数据本身开放程度/ a_{34} 论文发表要求/ a_{49} 不考虑发表/ a_{56} 毕业要求/ a_{69} 有选择的发表/ a_{72} 公开一部分

范畴化	概念化
A ₁₀ 感知利益	a ₁₈ 数据支撑论点/a ₂₈ 数据成果意义/a ₃₈ 人力物力付出/a ₄₃ 决策价值参考/a ₅₈ 同行交流/a ₆₇ 数据揭示角度不同
A ₁₁ 权益归属	a ₃₅ 团队实验成果/a ₃₆ 收集过程艰辛/a ₅₃ 自己的数据就发表/a ₇₀ 非公开数据公开较难/a ₇₁ 合作获得的数据
A ₁₂ 权益保护	a ₁₄ 更好保护数据/a ₃₇ 劳动成果认可/a ₆₁ 预注册/a ₇₇ 征求他人同意
A ₁₃ 道德要求	a ₃₉ 重复验证/a ₅₅ 资源共享实现/a ₅₇ 前人共享树立了榜样/a ₆₀ 抵制学术作假
A ₁₄ 数据价值	a ₁₃ 研究数据价值/a ₄₂ 研究质量/a ₅₂ 原始数据不能反映问题/a ₅₁ 数据量大且乱/a ₆₄ 数据使用价值低/a ₇₃ 数据质量控制/a ₇₆ 个人能力

2.2.2 主轴编码

通过分析各个概念层面的内在联系,最终将重复概念进行整合,最终得到14个范畴。通过分析范畴之间的联系,对14个范畴进行归纳、聚类和组合,形成个3主范畴,分别是数据管理能力、数据权益管理和数据伦理,如表5所示。

表5 科研数据共享影响因素主轴编码

主范畴	开放式编码	内 涵
数据伦理	数据来源	公开数据 指的是通过公共渠道下载、获取得到的数据内容 非公开数据 通过实验、调查、索取、购买等途径得到的未发布的数据
	道德要求	指的是在发表自己数据、重用他人数据、共享自己数据等过程中应遵循的道德准则
	共享行为习惯	指的是共享数据的主体通常在共享数据时存在的那些习惯,如不发表、发表部分等
数据管理能力	数据处理软件	可以帮助使用者进行数据处理的各种工具及软件
	数据处理方法	指的是对数据进行的各种操作方法
	基本素养	指利用、发现、解释数据等过程中具备的基本素养
	数据管理策略	指通过遵循发布政策、使用平台等基础设施、积极合作等策略对数据进行管理
	数据管理过程	对数据进行检索、组织、存储、维护、利用等操作
	数据处理能力	指在获取、处理数据过程中具备的能力
数据权益	感知利益	认为数据共享可能会给自己带来好处
	数据价值	数据作为资产对社会或个人产生积极作用的性质
	权益保护	是对数据管理过程中各利益相关者所代表权益的保护
	权益归属	确定数据的产权属于谁、为谁所用、如何使用等问题

2.2.3 选择性编码

通过分析表4的内容,确定了未来科研人员数据管理服务需求的影响因素的核心范畴为数据权益管理;基于对这一核心范畴的分析,又由于数据权益管理影响着数据权益管理技能的服务效果,确定权益管理需求是伦理需求的前提条件。因此,确定了“数据权益管理”这一核心范畴。

2.2.4 理论饱和度检验

在对剩余的1/3记录进行分析和处理时,仍能反映上述发现的主范畴。且所选取的检验内容中并未发现新的概念或范畴,在各主范畴内部也并未产生新的元素和内容。因此,本文得到的编码是饱和的。

3 理论模型构建与分析

通过上述所描述的三个步骤,得出 78 个初始概念、14 个范畴、3 个主范畴。本文在编码的基础上构建了未来科研人员数据管理服务需求影响因素分析模型(图 1),便于范畴的量化与理解,并得出重要结论。

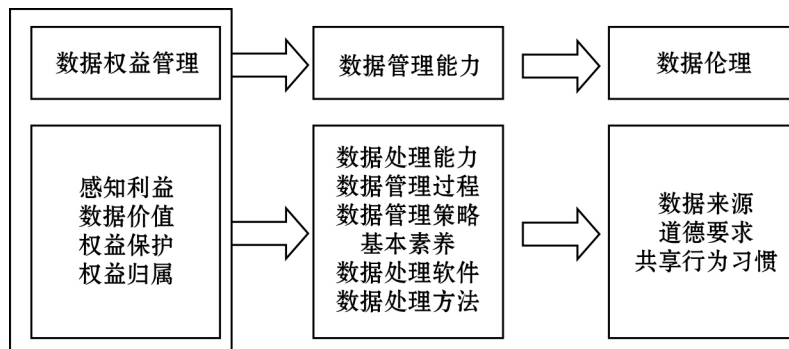


图 1 未来科研人员数据管理服务需求影响因素模型

在图 1 中,数据权益管理是核心影响因素,这一需求反映了研究生对于自身利益的诉求及数据责任的担忧。数据管理能力反映了研究生对于数据管理基础服务的需求,包含了对数据管理过程中所包含的服务要素。数据伦理反映了研究生在管理数据过程中应遵循的要求,技能服务需求和伦理需求是数据权益管理得到满足下的更高层次需求,而伦理需求则是权益管理需求和技能服务成熟的条件下被唤醒的需求。

3.1 数据权益管理

数据权益管理的影响因素包括感知利益、数据价值、权益保护、权益归属四个方面。大多数研究生认为数据共享会给自己带来诸多利益,例如数据引用、工作机会等,即感知到的利益。这是很多研究生通过数据管理等工作,提高数据可见度的动机。其次,数据的价值也需要得到评估和认可。访谈发现,很多研究生认为自己所产生的数据质量不高,没有共享的必要性,缺乏对生产数据的自信。同时,他们怀疑自己无法保证数据的共享价值。也担忧共享这些质量差的数据,给自己带来一些困扰。

针对数据权益保护和归属问题的关注度也很高。数据的类型通常有原始数据、衍生数据和科研数据三种。数据在流转过程中,经历了不同的产权主体^[50],如何界定不同类型的数据在不同阶段的产权是值得探讨的问题。而对这一问题的关注,也反映在了访谈结果中。访谈发现,研究生通常会有对自身数据被侵权的担忧及侵犯他人数据产权的担忧,渴望自己在共享数据的同时,掌握对数据的控制权。

3.2 数据管理能力

数据管理能力的影响因素包括数据处理能力、数据管理过程、数据管理策略、基本素养、数据处理软件、数据处理方法 6 个方面。其中,数据处理能力指的是研究生在检索数据、描述数据等方面的能力。数据管理过程体现了研究生对数据管理流程及所包含要素的掌握程度。数据管理策略主要指的是研究生借助政策、基础设施、工具等来辅助数据管理工作的方法。基本素养则主要是指研究生对数据检索、数据建模、数据报告撰写、利用数据解决研究中问题的能力。数据处理软件包含了研究生自身对数据工具的了解掌握情况及面临数据管理问题的能力。数据处理方法包括定性、定量处理方法两个方面。对定性数据来说,如何凭借分析者的经验、直觉来分析、对比、挖掘数据内容等方法的掌握非常重要。对定量数据来说,如何利用统计数据反映研究问题,并能解决研究问题等方法的使用非常重要。

数据管理能力需求,既反映了研究生对目前图书馆数据管理服务的满意程度,也包含了图书馆部署数据管理服务过程中应包含的服务要素。结果显示,数据来源多为数据集平台,而这些数据集往往为已被描述过的数据。大多数研究生不清楚什么是元数据,对数据描述工作的内容也知道得较少;而从研究生的回答得知,数据清洗、描述等的工作对最终的实验结果的效果有着重要的意义。此外,研究生数据处理工作要求研究生掌握软件操作和数据建模。

数据管理的有效性,将帮助研究生提高数据的使用效率,也是推动研究生共享数据的手段。对于有效的科研数据管理,首先要明确数据存缴层,即“人”的问题——数据的创建者/提供者^[34]。图书馆的数据管理服务虽然已经得到了各个方面的认可,但从上述资料可知,绝大多数研究生对数据管理能力仍较差,甚至很多人不清楚数据管理的流程。在问及数据处理的过程时,他们很难完整地叙述整个过程,回答内容多集中在数据处理的某一环节上。其次,发现大多研究生通常从数据存储平台中获取数据,却从来没有在这些数据集平台上存储过自己的数

据。同时，研究生对数据如何使用、去哪存储，没有明确的目标，缺乏规划。

3.3 数据伦理

数据伦理的影响因素包括数据来源、道德要求、共享行为习惯。数据来源通常包含公共渠道和非公共渠道。对于公共渠道数据的获取，通常要遵循数据发布者的要求及利用规范。对于非公共渠道获取到的数据，如实验室数据、企业内部数据等，使用和发布的要求通常会不同。访谈发现，研究生对于这种类型数据的获取和发布，往往会更加小心。

数据在管理过程中，往往会涉及很多的道德因素，例如数据的发布要求、数据质量的保证、学术不端行为的抵制等。道德标准在数据采集、发布、重用等步骤中起着积极指导的作用。访谈发现，研究生关注的道德因素集中表现在学术造假行为抵制、政策发布要求遵循、社会价值的实现等。此外，经过长时间的数据管理工作实践，科研人员会养成自己的行为习惯，如果正面影响数据管理工作，数据管理工作的效率将会大大提升，数据共享的价值也能得到更大化的发挥。

4 面向未来的图书馆的数据管理服务设计

基于以上步骤和分析，本文构建了图2所示的数据管理服务框架，包含了服务层级、服务内容两大方面。

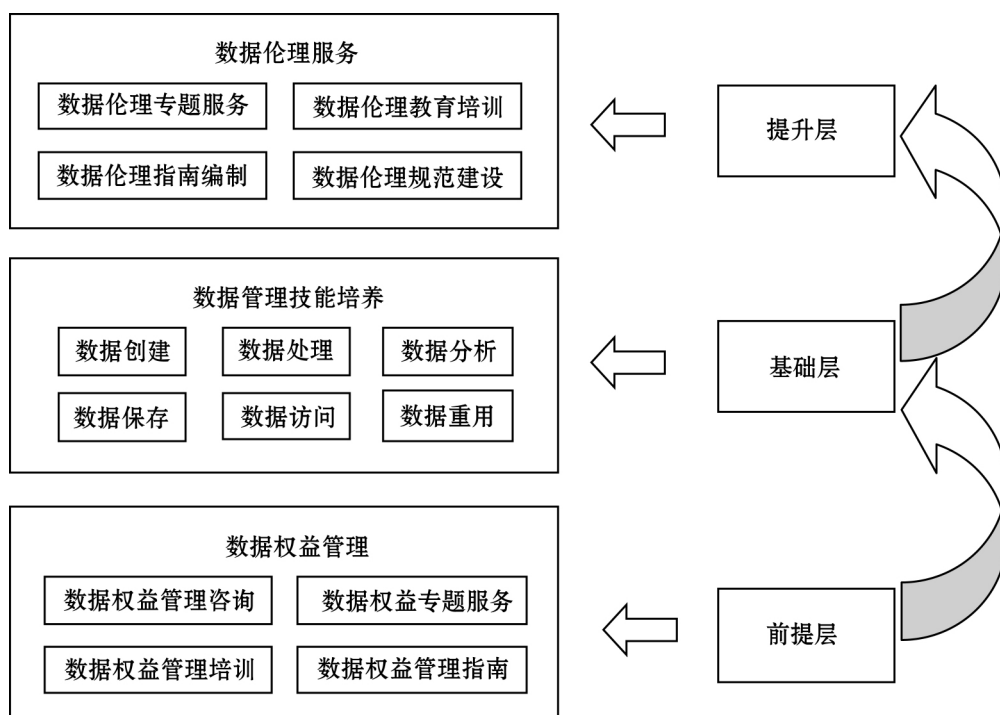


图2 数据管理服务框架

4.1 服务层级

根据马斯洛需求层次理论，本研究认为，对数据管理服务来说，也存在这样的层级。即数据管理的核心是数据权益，首要问题是明确科研数据管理的各利益方^[51]。在拥有了共同的目标和认可的价值观的基础上，才能平衡好各方的利益，保障数据管理技能培养工作的进行。

4.1.1 前提层

所有的数据都具有价值，都有被利用的可能。但数据的主权问题带给数据使用者多种困扰，如果处理不当，数据在管理过程中就会出现各种各样的问题，例如数据的利用边界权限不清、数据产权纠纷、数据的开放共享工作难以持续推动等。本研究在设计未来图书馆数据管理服务的框架时，将数据权益管理作为图书馆进行数据管理服务时的前提，要求图书馆能有专门的人员清楚地梳理数据权益管理存在哪几种类别的问题。

4.1.2 基础层

目前，数据管理服务的框架和要素在实践中不断发展，但结果显示，很多研究生的数据管理能力仍较差。本文认为，数据管理服务不仅注重服务的全面性，更要让研究人员真切地感受到服务的内容，更加注重服务的效果。

4.1.3 提升层

上述调研发现,研究生目前非常关注与数据管理法律、政策及伦理问题,尤其是如何利用伦理判断解决遇到的伦理难题。例如,数据引用的合理性、数据获取的正当性、数据发表的合规性。在提升层,图书馆的数据管理服务集中解决数据伦理规则的歧义,明晰数据伦理的规范,帮助科研人员养成合乎社会发展的行为习惯。

4.2 服务内容

本文通过梳理图书馆数据管理服务综述,分析上述访谈调查结果,对图书馆管理服务的内容进行了梳理,具体内容见表6。

表6 数据管理服务要点内容说明

数据管理服务类型	服务内容	服务要点
数据权益管理服务	数据权益管理咨询	数据管理咨询平台、数据权益管理专题服务等
	数据权益专题服务	数据权益问题动态监测、数据权益问题类别梳理等
	数据权益管理培训	用户数据权益教育、人才队伍素质培养等
	数据权益管理指南	数据边界标准、数据使用协议规范、数据权益分配机制及实施细节整理等
数据管理技能服务	数据创建	数据管理计划、数据描述方法、数据检索素养、元数据管理培训等
	数据处理	数据格式规范化、数据软件提供、数据科学课程教育等
	数据分析	数据分析的方法培训、数据论文撰写培训、数据分析工具提供等
	数据保存	数据保存培训、数据保存格式、数据长期保存、数据存储平台搭建等
	数据访问	数据访问的要求、数据访问限制、数据安全防范、数据访问培训等
	数据重用	数据重用机制、数据引用标准、数据关联、数据发现、数据标准管理等
数据伦理服务	数据伦理专题服务	数据伦理动态监测、数据伦理问题类别梳理等
	数据伦理指南服务	数据伦理参考手册、数据伦理判断指南等
	数据伦理教育培训	数据素养教育、数据伦理专题培训等
	数据伦理规范建设	数据伦理社群文化建设、数据伦理体系建设、用户行为习惯培养等

图书馆的数据管理服务设计要求通过日常工作的总结,梳理在数据权益管理服务、数据管理技能服务、数据伦理服务过程中的规则内容。其次,服务效果的评估也是未来图书馆数据管理服务需要提升的地方,特别是对培训工作专业性的提升及培训范围的扩大。

5 图书馆数据管理服务建议

5.1 数据权益管理精细化

数据权益管理是数据管理的核心和前提,数据权益问题解决措施的好坏,关系到各利益相关者是否认可相同的数据价值体系,是否能有效解决数据利用的壁垒,通畅地推动数据的循环利用。目前,数据权益管理讨论集中在数据权益主体框架和关系厘清、数据产权的归属与界定、授权许可方式等方面,研究主题通常侧重于某个问题,关于数据权益管理更全面、更深入的解决办法等的研究则比较缺乏。因此,未来图书馆的服务内容应精细化,详细地梳理数据权益管理过程中的主体、主体应承担的责任、主体享有的权益、主体之间的互动关系、数据产权归属及使用界限界定、数据授权许可协议标准等,为用户提供数据权益管理专题服务、数据权益管理培训服务、搭建数据权益管理的咨询平台和讨论组、为用户编撰数据权益管理手册等服务内容。

5.2 提高数据管理能力的服务效果

数据管理服务的内容经过长时间的发展,已经趋向成熟。大多数研究人员没有接受过科研数据管理方面的任何培训,比如数据管理计划、元数据或文件命名。然而,大多数研究人员欢迎科研数据管理不同方面的正式培训^[52]。因此,在开展数据管理服务时,在关注数据管理服务内容丰富的同时,也要注重数据管理能力的服务效果。

5.2.1 重视培训工作的作用

图书馆为用户提供了数据管理的资源和工具,耗费了大量的精力制定了元数据标准,用户却不知道如何使用,去哪使用。而数据管理培训工作针对性强,形式多样,可有效地帮助图书馆用户提升数据管理的能力。因此,图

图书馆数据管理培训工作值得重视。此外,在设计图书馆的管理培训工作时,图书馆应对不同的用户群体采用多方位的用户调查,之后根据不同的用户群体设计不同层次培训内容。

对学生来说,图书馆设置数据科学课程是普及数据管理基础知识、提高数据技能的有效方式。虽然一些高校的图书馆已经设置了专门的数据科学培训班、数据管理主题讲座,但涉及的学生群体不广,参与度低,难以达到普及的效果。如若图书馆与各学科院系积极合作,设置与数据管理有关的必修课和选修课,可在培养学生利用数据、管理数据能力和技能的同时,也能更近距离地与这一类型的用户接触。进而持续关注早期研究人员的研究需求,实时更新部署新型数据服务以契合科学界的未来发展需求。

刚踏入研究岗位的研究人员,对数据管理工作有一定的了解,但却未参与过数据管理的工作。对于这一类型的用户来说,数据管理工作的流程、工具、要求等方面的普及教育就显得很有必要。图书馆可以主动地与本单位负责人事的人员联系,定期为这些刚进入研究岗位的人员提供专题培训,并设置讨论组,实时解答研究人员在数据管理过程中的难题。

而对资深的研究人员来说,已经对数据管理工作有了较深的了解,但面对面的培训工作几乎是没时间参与的。图书馆在面向这一类型用户服务时,应秉承“参与为主,指导为辅”的原则,积极参与并了解科研人员的数据管理工作难点,主动和科研人员一起承担数据管理工作,有针对性地解决这类科研人员的数据管理工作难题。

5.2.2 提高数据管理服务的有效性

图书馆的服务最终应是面向用户需求,切实地帮助用户解决数据管理难题。因此,图书馆在部署数据管理服务过程中,应更加关注最终的服务效果。现实中,图书馆的用户通常面临着巨大的科研压力,很难主动地发现图书馆的服务内容,也缺乏时间和精力了解图书馆服务的内容。因此,针对这一问题,图书馆应与用户建立连接关系,多利用社会科学的方法,挖掘用户管理数据面临的问题和急需解决的难题。

5.3 逐渐开展数据伦理服务

数据伦理的目标是为用户提供伦理判断指南,是帮助数据管理工作有序运作的保障。科研人员对数据伦理内容的关注度随着数据管理实践的发展不断攀升。尤其是随着各领域数据量的不断积累,数据资产性质的确立,如何平衡数据隐私保护和数据共享利用之间的关系,成为了科研人员面临的数据管理难题之一。因此,图书馆应开始重视对数据伦理的服务内容,结合现有的数据管理服务工作经验,帮助用户明晰数据管理工作遇到的伦理问题。并通过数据伦理专题服务、数据伦理咨询服务、数据伦理教育培训、数据伦理规范建设等服务内容反馈给研究人员,使其不断地了解在数据管理过程中,哪些是值得参照的标准,哪些是需要摒弃的陋习,进而养成有益于数据管理良好运作的行为习惯。

6 结语

图书馆的数据管理服务经过长时间的发展与实践,其服务内容和形式得到了充实。但本文的访谈结果显示,很多研究生的数据管理能力存在不足之处,未来的图书馆服务还存在值得探索的空间。因此,图书馆在设计未来数据管理服务时,可将数据权益管理作为核心和前提,持续关注数据管理的服务效果,并将用户数据管理技能的提升作为服务目标。同时,在此基础上,关注图书馆用户的数据伦理需求,逐渐开展数据伦理服务。

注释

[1] 顾立平,包秦雯. 国外开放科研数据的激励政策[J]. 图书馆论坛, 2019(8): 138-145.

[2] 沈玖玖,王志远,戴家武,等. 基于扎根理论的科研数据需求及影响因素分析[J]. 情报杂志, 2019(4): 175-180, 160.

[3] NSF. Revolutionizing Science and Engineering Through Cyber Infrastructure [EB/OL]. [2020-03-16]. <https://www.nsf.gov/cise/sci/reports/atkins.pdf>.

[4] NSF. Dissemination and Sharing of Research Results [EB/OL]. [2019-10-02]. <https://www.nsf.gov/bfa/dias/policy/dmp.jsp>.

[5] UKRI. UK Research and Innovation: Data Policy [EB/OL]. [2019-10-02]. <https://www.ukri.org/funding/information-for-award-holders/data-policy/>.

[6] 新华社. 国务院办公厅关于印发科学数据管理办法的通知[EB/OL]. [2019-10-02]. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.

[7] European Commission. European Open Science Cloud (EOSC) [EB/OL]. [2019-10-02]. <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>.

[8] OpenAIRE. About OpenAIRE [EB/OL]. [2019-10-02]. <https://www.openaire.eu/mission-and-vision>.

[9] arXiv. About arXiv [EB/OL]. [2019-10-02]. <https://arxiv.org/>.

- [10]Dataverse. About the Project [EB/OL]. [2019-10-02]. <https://dataverse.org/about>.
- [11]ANDS. About ANDS [EB/OL]. [2019-10-02]. <https://www.ands.org.au/>.
- [12]SINET. About Science Information Network 5 [EB/OL]. [2019-10-02]. <https://www.sinet.ad.jp/en/top-en>.
- [13]CNIC. China Science and Technology Cloud Reported by China Daily [EB/OL]. [2019-10-02]. http://english.cnic.cas.cn/news/events/201805/t20180510_192384.html.
- [14]地球大数据科学工程. 地球大数据科学工程先导专项概况[EB/OL]. [2019-10-02]. <http://www.casearth.com/>.
- [15]Digital Belt and Road. About DBAR [EB/OL]. [2019-10-02]. <http://dbelroad.org/index.html>.
- [16]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 (8): 159-167.
- [17]Springer Nature Advancing Discovery: How We Are Helping the Research Community to Advance Discovery [EB/OL]. [2019-10-02]. <https://www.springernature.com/gp/advancing-discovery>.
- [18]李晓辉. 图书馆科研数据管理与服务模式探讨[J]. 中国图书馆学报, 2011 (5): 46-52.
- [19]马建玲, 祝忠明, 王楠, 等. 美国高校图书馆参与研究数据管理服务研究[J]. 图书情报工作, 2012 (21): 77-82, 142.
- [20]李丹丹, 吴振新. 研究数据管理服务综述[J]. 图书馆学研究, 2012 (9): 54-59, 69.
- [21]王学勤, Stout A, Silver H. 建立数据驱动的 e-Science 图书馆服务: 机遇和挑战[J]. 图书情报工作, 2011 (13): 80-83.
- [22]洪程. 国外科学数据服务现状研究[J]. 图书馆杂志, 2012 (10): 31-34.
- [23]费若云. 图书馆实施分阶段科研数据管理服务中的权益问题初探[J]. 山东社会科学, 2016 (S1): 615-617.
- [24]邹中才, 顾立平, 张闪闪, 等. 图书馆实施分阶段科研数据管理服务中的权益问题初探[J]. 情报理论与实践, 2016 (1): 64-69.
- [25]顾立平. 科学数据权益分析的基本框架[J]. 图书情报知识, 2014 (1): 34-51.
- [26]张闪闪, 顾立平, 盖晓良. 国外信息服务机构的数据管理政策调研与分析[J]. 图书情报知识, 2015 (5): 99-109.
- [27]刘兹恒, 曾丽莹. 我国高校科研数据管理与共享平台调研与比较分析[J]. 情报资料工作, 2017 (6): 90-95.
- [28]王辉, Witt M, 冀天芳. 普渡大学研究仓储及其支持的科学数据管理服务[J]. 现代图书情报技术, 2015 (1): 9-16.
- [29]魏来, 郑华敏. 国内外数据馆员能力要求比较研究[J]. 图书情报工作, 2018 (10): 18-24.
- [30]陈媛媛, 柯平. 大学图书馆数据馆员职业技能发展研究[J]. 情报资料工作, 2018 (2): 97-102.
- [31]郝媛玲. 关于数据服务模式与数据馆员新角色的思考[J]. 图书馆学研究, 2018 (4): 80-85.
- [32]胡绍君. 面向科研数据管理的高校学科馆员能力建设研究[J]. 图书情报工作, 2016 (22): 74-81.
- [33]胡卉, 吴鸣, 陈秀娟. 英美高校图书馆数据素养教育研究[J]. 图书与情报, 2016 (1): 62-69.
- [34]鄂丽君. 国外大学图书馆的科研数据管理教育[J]. 情报资料工作, 2014 (1): 101-105.
- [35]蔚海燕, 卫军朝. 国外高校数据监管项目的调研与分析[J]. 图书情报工作, 2014 (22): 38-47.
- [36]黄如花, 邱春艳. 图书馆参与科学数据管理中的元数据应用实践研究[J]. 图书与情报, 2014 (5): 65-69.
- [37]党洪莉, 谭海兵. 基于 DMM 的数据管理成熟度模型及在服务评估中的应用[J]. 现代情报, 2017 (9): 118-121.
- [38]叶兰. 研究数据管理能力成熟度模型评析[J]. 图书情报知识, 2015 (2): 115-123.
- [39]王思明. 美国高校图书馆数据管理计划服务及启示[J]. 数字图书馆论坛, 2018 (12): 34-40.
- [40]王丹丹. 数据管理计划在图书馆科学数据管理服务中的应用[J]. 图书情报工作, 2018 (1): 76-81.
- [41]王静, 都平平, 郭太敏, 等. 基于四螺旋模型的高校图书馆科研数据知识服务风险管理研究[J]. 图书馆, 2015 (3): 79-81, 85.
- [42]李志芳. 国外 iSchool 联盟高校图书馆科研数据管理调查[J]. 图书馆论坛, 2017 (10): 121-130.
- [43]孟祥保, 叶兰, 常娥. 高校图书馆科研数据联盟建设策略——以荷兰 3TU. Datacentrum 为例[J]. 图书情报工作, 2015 (2): 31-37.
- [44]周雷, 刘利永. 德国 RISE-DE 科研数据管理服务自评估模型的研究[J]. 图书馆学研究, 2019 (21): 38-47, 9.
- [45]柴会明, 张立彬, 赵雅洁. 国内图书馆科学数据研究述评[J]. 图书情报工作, 2019 (7): 116-126.
- [46]张培风, 张连分. 全球科研范式变革下的图书馆科学数据管理服务创新——基于数据管理生命周期的视角[J]. 图书馆理论与实践, 2019 (5): 39-48.
- [47]杨乐, 颜石磊, 李洪波. 科研数据生命周期研究和数据知识库理论架构[J]. 图书情报工作, 2019 (1): 91-97.
- [48]李梅. 大学图书馆的研究数据服务创新与数据馆员新角色——英国爱丁堡大学范例研究[J]. 图书与情报, 2019 (3): 122-130.
- [49]许磊, 夏翠娟. 第三代图书馆服务平台的元数据管理——以 FOLIO 的 Codex 方案为例[J]. 中国图书馆学报, 2020 (1): 99-113.
- [50]顾立平, 樊舒. 数据权属和使用边界研究[J]. 科研信息化技术与应用, 2018 (3): 34-39.
- [51]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 (6): 43-58.
- [52]Melero R, Molina C. Researchers' Attitudes and Perceptions Towards Data Sharing and Data Reuse in the Field of Food Science and Technology [J]. Learned Publishing, 2020, 33 (2): 163-179.

刘金亚 中国科学院文献情报中心, 中国科学院大学经济与管理学院图书情报与档案管理系硕士在读。研究方向: 科研数据管理、数据伦理。通讯作者。

顾立平 中国科学院文献情报中心研究员, 博士。研究方向: 科技信息政策、科研数据管理。