



高校图书馆数据治理框架构建

——复旦大学图书馆的实践与思考

薛崧 张计龙* 苗青 任磊 毛梦丹 赵敏 张耀

摘要 在教育数字化转型和人工智能驱动的科学范式变革的背景下,数据治理已经成为高校图书馆提升数据管理能力、应对现代化转型发展的必然选择。以数据治理理论和相关国家标准为指导,从高校图书馆数据工作的现实问题出发,构建了包含战略引领、治理内容、治理过程、平台工具、治理规范和组织架构六个要素的高校图书馆数据治理框架,并以复旦大学图书馆实践为例,探讨数据治理的关键方法、挑战和应对建议,以期为高校图书馆的数据治理、数据赋能的实践提供参考。

关键词 高校图书馆 数据治理 数据标准化 数据治理框架

分类号 G250.7

DOI 10.16603/j.issn1002-1027.2025.02.008

引用本文格式 薛崧,张计龙,苗青,等.高校图书馆数据治理框架构建——复旦大学图书馆的实践与思考[J].大学图书馆学报,2025,43(2):84-94.

1 引言

随着教育数字化转型的加速推进,数据治理已成为一个需要关注和解决的关键问题。美国高等教育技术协会(EDUCAUSE)在其2023年报告中强调,有效的数据治理对实现高等教育数字化转型至关重要^[1-2]。教育数字化转型也为高校图书馆带来了新的挑战与机遇。2022年,教育部高等学校图书馆情报工作指导委员会联合北京大学图书馆发布的《大学图书馆现代化指南针报告》^[3]提出,数据资源与传统图书馆信息资源的互补和融合发展,是推动大学图书馆服务创新的坚实保障和动力来源。该报告倡导高校图书馆应基于现有数字资源进一步挖掘、构建数据资源、数据资产和数据要素,以实现数据资源的增值服务^[4]。因此,高校图书馆开展数据治理不仅是适应教育数字化转型的必然选择,也是实现图书馆现代化转型和支撑学校高质量发展的关键所在。

数据治理是通过对战略规划、组织制度、技术工具和标准规范等的综合运用,对机构的数据管理活动进行评估、指导和监督,以实现数据资产价值的最

大化^[5-7]。

目前,图书馆领域多从宏观和理论层面研究数据治理,研究内容涵盖图书馆数据治理框架、成熟度模型和评价体系^[6,8-9]、科研数据治理^[10-11]、知识管理中的数据治理^[12]、基于数据中台的数据治理体系^[13]等。在此基础上,更多的学者开始探讨图书馆数据治理中的具体问题。李琳等提出高校图书馆协同参与高校整体数据治理的概念框架^[14]。杨新涯等提出包括六种类型的智慧图书馆全数据体系以及建设智慧图书馆的五种关键数据能力^[15]。陈飞等根据PDCA模型设计了基于业务目标的数据治理流程^[16]。卢小宾等提出包含概念术语定义、格式编码、框架体系和操作规范等的智慧图书馆数据标准规范体系^[17]。刘倩倩等认为与大模型训练、微调和应用开发相关的数据治理需求已逐步显现,是实现大模型垂域应用效果的关键^[18]。

在实践层面,图书馆普遍关注智慧服务的形式、内容以及效果,忽略数据基础工程的构建^[15],把数据治理作为有意识、体系化工作的图书馆很少。复旦大学图书馆于2023年开展的一项关于高校图书

* 通讯作者:张计龙,ORCID:0000-0003-1970-2881,邮箱:jlzhd@fudan.edu.cn。



馆数据工作现状及治理需求的访谈显示,高校图书馆存在一定程度的数据危机,其数据管理和利用能力跟不上对数据价值的期待^[19]。具体表现为:(1)缺乏数字化转型战略和数据战略、数据治理的顶层设计以及可操作的数据治理方法;(2)传统的图书馆资源组织规范难以实现多源异构数据的融合,加剧数据孤岛问题;(3)数据标准化、质量、安全与合规性,以及高校和图书馆层面的管理保障机制存在诸多问题,影响了数据对创新应用和管理决策的支持;(4)缺乏必要的技术能力和适配的平台工具,使数据的采集、存储、整合和服务难以满足新的业务需求;(5)缺乏既懂图书馆业务又精通数据和技术的复合型人才;(6)存在校内跨部门及行业内外数据协作需求,需规范和监督数据的共享使用。

数据治理重在实践。复旦大学于2020年启动智慧图书馆建设,同年成立“复旦—阿法迪共建智慧图书馆学研究中心”(以下简称“智图研究中心”),旨在通过产学研用合作机制促进智慧图书馆的研究、实践、成果转化和跨学科人才的培养。2022年8月,复旦大学智慧图书馆数字基座正式上线发布,作为智慧图书馆的数字基础设施,其核心任务是支撑高校图书馆数据治理、提升数据能力。依托数字基座,复旦大学图书馆逐步解决了数据孤岛和数据标准化等问题,并形成了一套系统化的数据治理方法。本文以数据治理理论和相关国家标准为指导,在教育数字化转型发展背景下,结合复旦大学图书馆近年来的数据治理实践,探讨高校图书馆数据治理框架的主要构成要素以及关键实践,以期为其他高校图书馆的数据治理、数据赋能的实践提供参考。

2 高校图书馆数据治理框架

国际数据管理协会(Data Management Association,DAMA)、国际数据治理研究所(Data Governance Institution,DGI),国内外标准组织、信息技术领域的专业机构和咨询公司等从不同的视角,提出了功能和适用性各有侧重的通用数据治理框架和模型,为数据治理的实践提供了理论指导。其内容构成要素主要包括范围、过程、评估、角色、原则、职能^[7],以及为数据治理实施提供保障的环境要素^[20]。我国现行的国家标准为企事业单位的数据治理体系建设、实施及效能评估提供了更为具体的指导。例如,《GB/T 34960.5-2018 信息技术服务治

理第5部分:数据治理规范》提出了由顶层设计、数据治理环境、数据治理域和数据治理过程组成的数据治理框架,《GB/T 36073-2018 数据管理能力成熟度评估模型》则定义了8个数据管理能力域及成熟度评估方法和流程,《GB/T 44109-2024 信息技术 大数据数据治理实施指南》和《GB/T 42450-2023 信息技术 大数据 数据资源规划》在上述标准基础上提出了大数据环境下实施数据治理的方法和路径(下文涉及以上国家标准,均以国标代码简称)。

在高校图书馆的数据治理实践中,可以以GB/T 34960.5-2018的总体框架为指导,参考通用数据治理框架的构成要素及相关国家标准,结合高校图书馆的发展战略和当前数据工作面临的挑战,制定并细化数据治理框架的具体内容和要求。高校图书馆作为学校的二级单位,其数据治理的顶层设计和制度体系应遵循学校的相关规定,在实施过程中需重点关注影响数据治理工作成效的关键要素^[6],包括战略与目标、活动与内容、流程与评估、技术与工具、标准与规范、角色与职责。相应地,一套完整的高校图书馆数据治理框架可从战略引领、治理内容、治理过程、平台工具、治理规范和组织架构六个方面来构建,如图1所示。与通用数据治理框架相比,该框架提出了在教育数字化转型发展背景下,高校图书馆以支撑教学科研为核心目标,针对学术数据资源管理的具体问题和需求,通过规范化的流程、适配的IT能力和责权分明的团队架构,系统性地实施数据治理的方法。

2.1 战略引领

高校图书馆的数据治理工作应以国家和教育数字化转型战略为指引,在学校和图书馆的整体发展战略规划基础上制定数据战略、明确治理的目标和任务,并依此推进一系列数据工作,从而构建面向未来的高校图书馆高质量发展所需要的数据能力,以支撑整体战略目标的实现。

2.2 治理内容

高校图书馆的数据治理内容包括当前和未来一段时期里所需数据资源以及相关的数据管理和数据资源建设活动。

从规划视角来看,高校图书馆的数据资源应涵盖能够支撑转型发展需求的全域数据,通过“全面感知、数据全覆盖”,解决图书馆实体与数字、现实与虚拟、线上与线下的业务融合^[21]。



从高校图书馆数据工作现状来看,数据治理急需解决数据孤岛问题,并实现多源异构学术资源数据的汇聚和融合,从而形成规范、可用和持续增值的数据资源,以增强数据在多样化应用场景中的支撑能力。随着管理和创新实践的推进,学术资源之间基于内容的细粒度的关联,以及多模态、非结构化数据的管理,将成为数据深度治理的核心内容。

因此,高校图书馆应根据数据应用规划和业务优先级,确定数据治理的关键内容。可参考 GB/T 36073—2018 的数据管理能力域定义,以及 GB/T 42450—2023 的数据资源规划流程,以场景需求为牵引,分阶段构建数据管理体系和建设数据资源,打造“需要的全量”^[22]。

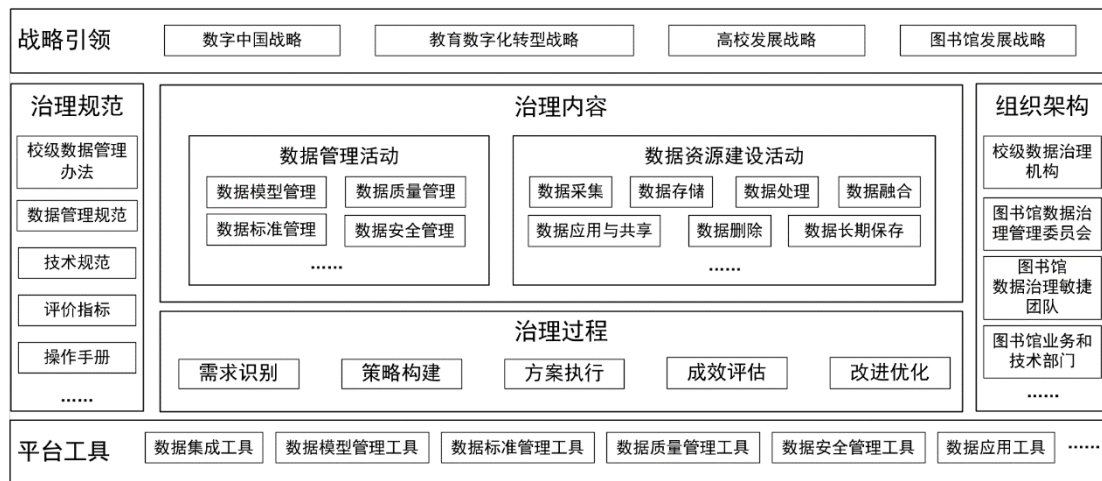


图 1 高校图书馆数据治理框架

2.3 治理过程

参考 GB/T 44109—2024,高校图书馆的数据治理过程是一个包括需求识别、策略制定、方案执行、成效评估以及持续的改进优化的闭环管理过程。

首先,需要从高校图书馆整体发展战略中识别出数据需求,明确数据治理的愿景、目标、任务,并制定实施策略和路线图,然后策划、构建和实施系统性数据治理方案。进一步地,通过监控和评估数据治理的执行情况和有效性,对方案、策略和方法等进行不断的优化和改进,从而持续提升治理成效。

2.4 平台工具

数据治理工作通常由一系列的平台工具组合作为组织、策略、规范、流程等落地的载体。参考 GB/T 34960. 5—2018、GB/T 44109—2024 提出的数据治理活动内容,数据治理平台工具通常涵盖数据集成工具、数据模型管理工具、数据标准管理工具、数据质量管理工具、数据安全治理工具,以及与数据共享、服务和分析相关的数据应用工具。

针对当前高校图书馆数据工作中存在的问题,平台工具应重点解决数据孤岛问题,实现多源异构学术资源数据的有效融合,并提升数据质量和合规

性。同时,应为馆员提供便捷、安全的数据应用工具,以增强其自主发现、获取和利用数据的能力。

2.5 治理规范

数据治理规范用于指导、规范高校图书馆的数据工作,同时确保数据治理相关软件或解决方案的研发、选择和评价有章可循。对应于数据治理的组织层次和决策次序,高校图书馆应在学校信息化管理制度和规范指导下制定数据治理相关的规范,包括但不限于数据管理办法,数据治理的技术规范、评价指标和操作手册等,相关的内容需尽可能借鉴、采用或引用已有的标准规范,并与之保持协调一致。例如,大数据、信息安全技术、信息技术服务、系统与软件工程等领域的国家标准,教育部发布的《教育基础数据》《教育系统人员基础数据》等行业标准,以及图书馆领域资源描述和组织规范等。

2.6 组织架构

建立清晰的数据治理组织架构,协调各方利益、规范各方行为,是数据治理成功实施的保障。基于目前高校图书馆的业务架构、部门设置和高校数据治理参与者角色,数据治理组织架构可分为 3 个层面:决策层、管理层和执行层。



决策层由校级数据治理机构和图书馆数据治理管理委员会组成,负责审批数据战略、数据治理规划等,并提供相关工作的资源保障、业务指导,协调馆内外跨部门事务等。

在数据治理工作起步阶段,管理层可由一个跨部门团队组成,负责制定数据治理方案并按计划部署实施,同时组织协调全馆数据相关工作的开展。

执行层是具体实施数据治理相关工作的人员。例如,负责维护文献资源数据的业务部门工作人员,信息系统建设人员等。管理层成员也可能是执行层的工作人员。

3 复旦大学图书馆数据治理的实践

2021年起,复旦大学图书馆启动数据治理工作,并参照前述框架着力推进六项关键举措,以确保数据治理理念转化为行动。

3.1 制定数据战略,明确数据治理的目标、关键任务和实施策略

复旦大学图书馆在“十四五”规划中提出服务提升、引领高度和基础保障三个方面的转型发展愿景,并制定了针对性的数据战略:多方汇聚资源,发挥馆藏优势,融通异构数据,构建能支撑智慧化管理和服务的数据基础设施,为师生教学科研、科研学术评价和学校治理决策提供数据资源和数智能力支持。

数据治理的总体目标包括:(1)增强全馆数据意识和数据思维,提升图书馆在转型发展中的突破和创新能力;(2)构建完整、规范、可用的图书馆数据资源体系,满足教学科研需求,并为精细化、智能化和个性化的管理与服务提供数据基础;(3)持续提升图书馆学术资源数据的质量和值,支撑可持续的业务转型和服务创新。

对数据资源和相关工作进行评估后,明确数据治理首要解决的是数据资源分散孤立等基础性问题,推进数据标准化;其次,是文献资源的数据化、知识化,并在各类学术资源数据之间建立语义关联和内容融合;同时,数据质量和数据安全是持续关注的重点内容。

相应的,遵循规划先行、按需治理的原则,在起步阶段重点解决数据工作中的基础性问题,然后逐步迭代、系统化地推进数据治理工作。

3.2 整合异构数据,构建标准统一和高质量的数据资源

针对文献资源元数据、本校学术成果数据、用户

数据、业务数据等存在的分散孤立、标准不一问题,复旦大学图书馆通过构建统一的数据模型,并在数据资源建设过程中针对不同来源和类型的数据实施分层、透明且差异化的处理流程,不仅有效整合了多源异构数据,还提升了数据的整体质量和可用性。

3.2.1 基于实体—关系的可扩展的数据模型

随着数据资源类型的不断丰富和智能化应用的持续增长,图书馆面临多种元数据格式的文献资源数据、多个来源的个人数据的整合,以及建立多种类型数据之间的关联关系的挑战。传统的图书馆资源组织规范在多源异构数据的融合上存在局限性,因此,复旦大学图书馆在 IFLA 图书馆参考模型^[23](IFLA Library Reference Model, IFLA LRM)的基础上,构建了一个统一且可扩展的实体—关系数据模型(Extendable Library Reference Model, eLRM),以适应数据资源类型的多样化和应用场景的创新需求。eLRM 遵循 IFLA LRM 的基础结构,将模型的实体从书目资源的相关对象扩展到更广泛的图书馆业务对象,除引入行为者(个人、团体)、文献资源(作品、内容表达、载体表现、单件)、地点、时间段等实体,还根据应用需求增加了场馆设施、事件(师生行为、业务操作)实体等,从而增强了数据的融合和关联分析能力。此外,扩展个人实体的属性,增加读者、学生、教师、责任者的附加属性,以融合从多渠道获得的不同特征的个人数据或以更细粒度方式记录数据;同时,补充了个人与文献资源之间的责任、收藏和捐赠关系等,以勾画实体间的关系图谱。

eLRM 不仅解决了目前高校图书馆异构数据融合和标准化问题,还为未来多源数据的集成打下了良好的基础,其特点包括:(1)通用性。采用国际图情领域广泛认可的概念模型,便于图书馆之间数据资源的共享和交换。(2)易扩展性。通过添加实体、属性和关系的方式,可以实现不同场景(如科研、教学、服务、管理等)中的数据以“插拔式”方式进行快速融合,数据组织方式能够灵活适应未来业务的变化。(3)灵活性。允许根据应用需求灵活定义实体属性和关系的颗粒度,使其能快速响应并精确反映业务需求的变化。通过持续的迭代,模型能与图书馆业务的发展保持同步,确保数据结构始终与当前业务实践相匹配。

3.2.2 分层、透明且差异化的数据资源建设流程

在数据资源建设过程中,图书馆利用数据集成工



具,按需从多个来源采集数据并存储为贴源数据,然后按既定规则加工成明细数据和业务对象数据。进一步地,这些数据可以被提炼为应用数据,以满足特定的场景需求。数据资源建设的总体流程如图2所示。

该流程的特点包括:(1)通过分层建设与存储,

将复杂的数据处理和数据冗余控制在前期,以便后续阶段能直接利用数据,提高其在不同场景的适用性和可复用性;(2)记录各层数据的血缘关系,增强处理流程的透明度,并支持数据溯源;(3)根据数据的类别特性和应用需求,实施差异化的处理策略。

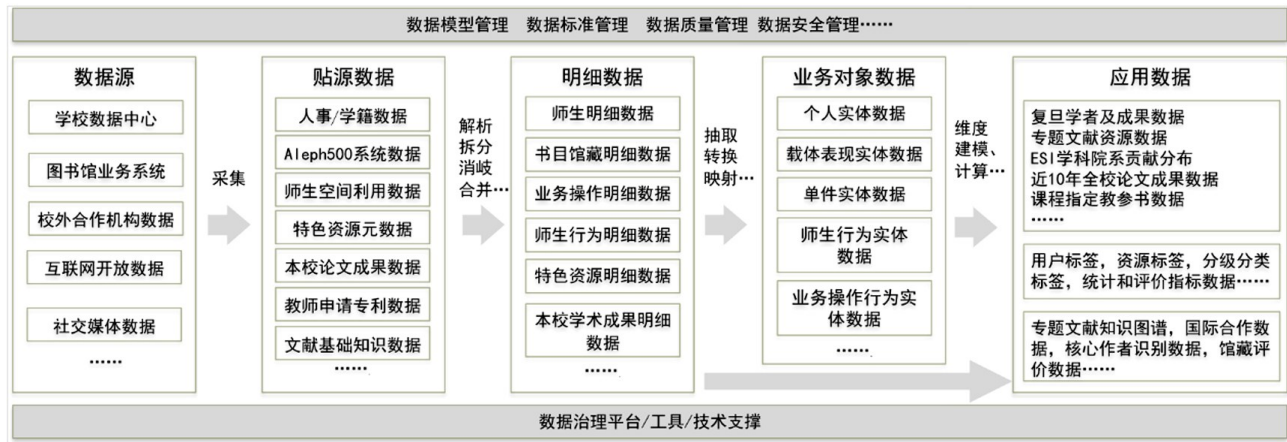


图2 数据资源建设总体流程

在处理来自多个渠道的具有不同特征的个人数据时,按图3所示的流程,采用自动加人工干预的方法,先对贴源数据执行解析、拆分等步骤,形成读者、学生、教师、学者和责任者的明细数据。然后,针对明细数据实施查重、消歧、关联、合并、抽取和转换等操作,

生成具有唯一ID的个人实体数据,其中每个字段的元数据记录了数据的来源、生成算法和映射规则。上述处理过程实现了个人数据的唯一性、一致性和可溯源性,便于整合师生在不同业务系统中的数据,形成更全面的用户画像。

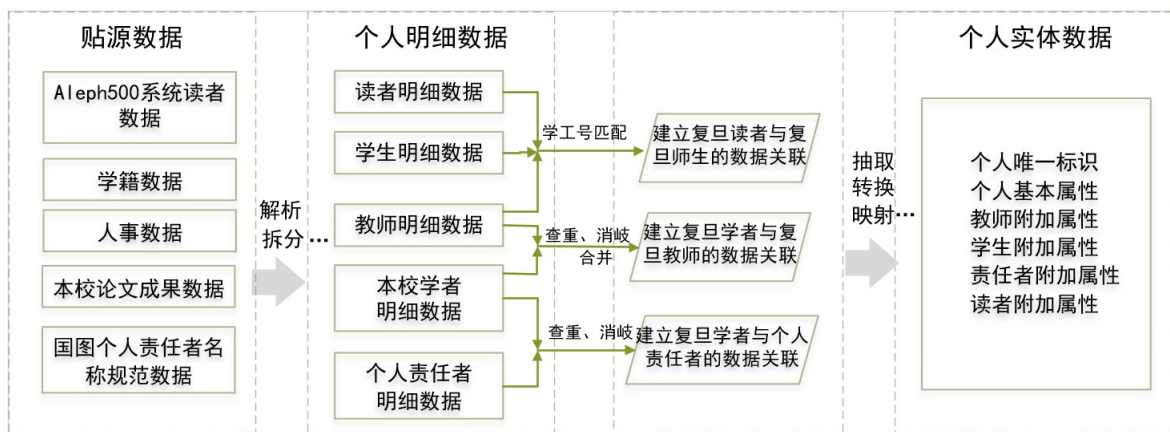


图3 多源个人数据的处理流程

针对多种格式的文献资源元数据,先按既定规则自动进行解析和拆分,形成维度丰富的明细数据;然后,建立 CNMARC、MARC21、DC、EAD 等与文献资源实体属性之间的映射关系和处理规则,并构建载体表现、单件实体数据及其与个人和其他实体

的关系数据。例如,为直观呈现重要捐赠人的捐赠资源,我们首先从 Aleph500 系统和 ArchivesSpace 非书资料管理系统分别采集 MARC 格式的书目数据和 EAD 格式的特藏资料元数据,以及单册记录中的标记信息(如签名、题赠、注释、印章等)。然后,根



据相关处理规则将两种格式的数据转换成载体表现和单件实体数据,并抽取标记信息以建立其与相关学者的关系。由此,将分散的文献资源通过个人实体有效地连接起来,可为师生提供网络化、体系化的馆藏资源推荐,以及提供重要学者、捐赠人与馆藏文献资源的关系网络。

针对多个服务系统的业务数据,则以个人和文献资源为中心进行重组,并补充行为、时空、状态及与其他数据的关联关系,形成一个全面、动态的数据视图。例如,从7个核心业务系统中采集读者行为的细粒度日志数据,在统一时间和空间等字段的数据标准后,整合形成读者行为明细数据,可为业务统计分析、阅读对账单、读者积分等应用快速提供基础数据支持。进一步地,该数据与师生个人信息结合,可提供更个性化的服务,并且,在智慧空间管理、采购决策支持、馆藏评价等多个场景中也将发挥重要的数据支撑作用。

3.3 以用促建、闭环管理,持续改进数据治理实施工作

复旦大学图书馆一方面积极拓展应用场景,以需求驱动数据资源建设;另一方面,根据使用反馈以及数据质量评估情况,对数据本身、数据模型、平台工具以及业务流程等进行改进,形成闭环的治理过程。

以“复旦大学教学参考资源服务系统”为例,图书馆与研究生院、信息化办公室合作,通过数据互通和流程协同,为选课学生提供课程指定教参的全文服务。

从试用情况看,数据质量对服务可用性和用户体验的影响较大。因此,图书馆重点从馆藏数据质量管理和业务流程优化两方面采取了相应改进措施。

首先,利用数据质量管理工具,在数据资源建设流程中实施了“事前准入检测、事中质量管控、事后质量评价和优化”的闭环数据质量管理机制,使数据不仅符合业务规范,而且符合用户的期望或预期用途^[24]。事前准入检测主要针对书商提供的外文电子书书目数据,在导入业务系统之前,通过预置规则对其进行合规性检测。事中管控则在数据处理环节实施。事后措施针对应用数据,根据场景需求定义数据质量量化评价模型,包括完整性、可访问性、一致性等评价维度以及相应的权重,每个维度设置多

个检测规则,如关键属性的空值检测和全文链接的有效性检测等。异常数据将通过数据血缘分析进行源系统定位和修改。馆员也可以对数据模型等进行优化和调整,以提升检索效率和效果。

其次,对图书馆教参资源管理流程进行重构,由学科馆员负责审核、确认和发布查重匹配后的数据,以确保数据的准确性。对于未匹配到馆藏的指定教参,采访馆员将优先启动电子图书快速采购流程。

这些措施共同提升了图书馆数据治理的成效,也为师生提供了高质量的教学支持服务。

3.4 构建平台工具体系,提升馆员数据管理和利用能力

2021年以来,复旦大学图书馆依托数字基座—数据中心,建设了一套包括数据集成、数据管理和数据应用在内的工具体系,这些工具相互关联、协同工作,构成一个综合性的数据治理平台,成为打破数据孤岛、提升数据质量和释放数据价值的关键基础设施。

数据集成工具能从多种数据库系统、API接口和数据文件中采集数据,并依据统一的数据模型和数据标准进行处理和分层存储,建设符合应用需求的数据资源。

数据管理工具覆盖数据模型、数据标准、数据质量和数据安全四个关键领域。数据模型管理工具为馆员提供了实体关系设计和管理的基础功能,而数据标准管理工具则用于数据标准的维护、发布等。在此基础上,建设数据质量管理工具,使馆员可以以低代码和可视化方式配置质量模型和检测规则,开展数据质量问题的识别、监测、评价和预警。此外,数据安全工具支持馆员在创建数据集时,根据数据模型中敏感属性的标识,设置数据分级、脱敏加密和访问控制等安全措施,以保护敏感数据。

在数据资源建设的同时,根据业务需求开展数据应用工具的建设。例如通过指标数据管理工具,集中管理全馆关键业务指标,为业务统计和分析提供统一的指标数据服务。又如,通过数据集市工具向馆员提供自助、可视化的数据集查询、创建、预览和下载服务,解决“数据找不到、读不懂”的问题。

从馆内使用反馈来看,这些工具不仅帮助馆员用新的思路和方法利用数据,而且降低了业务馆员使用数据的技术门槛,增强了他们对数据的理解和信任,有效提升了工作效率。



3.5 持续完善规范体系,建立有效的数据治理环境

数据治理实施之初,学校和图书馆层面均没有专门针对数据的管理办法和专门规范。为了确保数据治理工作有据、可行、可控,复旦大学图书馆参考教育大数据标准体系分类^[25]和上海教育数据标准体系框架^[26],围绕数据治理的重点工作和关键环节,逐步构建了包含管理、数据、技术、安全和平台工具等内容的规范文档。

其中,《复旦大学图书馆数据管理办法》明确馆层面的数据管理基本原则,以及各部门/岗位在数据生命周期中的职责分工。数据类规范用于指导数据标准化和数据质量管理,包括但不限于《数据中心实体属性描述规范》《个人实体的通用和附加属性以及数据标准》《载体表现的通用和附加属性以及数据标准》《单件的通用和附加属性以及数据标准》;技术类规范用于标准化数据采集和预处理流程,主要包括《数据中心数据集成技术规范》《数据中心个人数据处理规范》《多种编目格式与载体表现、单件的映射关系和处理规范》等;安全类规范对数据生命周期各个环节提出安全管理要求,并且通过《数字文献/数据分级安全管理和服务规范》,确保数据共享和使用的合法、合规;平台工具类规范对数据治理职能系统的开发、管理和使用进行约束,主要包括《数据中心整体架构与建设规范》《数据集市自助服务使用规范》《数据质量准入检测规则和质量评价模型自助配置手册》《指标数据自助配置手册》等。

目前实施的各类规范涵盖了从微观到宏观的多个管理层面,规范之间相互衔接和制约,协同保障了数据治理工作的有序推进。随着数据治理的持续深化,规范体系将不断发展和完善,以确保建立一个有效的数据治理环境。

3.6 立足全局,构建跨部门团队,促进数据治理的全面实施

复旦大学图书馆通过建立跨部门协作和动态调整机制,统筹推进全馆的数据工作。数据治理组织架构包括数据治理领导小组、数据治理工作小组以及图书馆各部门。此外,学校信息化办公室及校内外数据管理专家提供数据治理规划、策略和方法方面的专业指导。

数据治理领导小组由分管数据管理与技术、资源建设的馆领导组成,负责规划、组织、管理和保障等关键问题的决策和协调。

数据治理工作小组由业务、技术部门的骨干和智图研究中心的专职技术团队构成,承担数据治理的研究和实践工作。

小组的核心成员及其职责会根据阶段性任务的变化而进行动态调整。在数据治理的起步阶段,图书馆技术和业务部门负责人协同完成相关调研工作,并制定数据治理规划和实施方案,提交数据治理领导小组决策。在实施过程中,小组成员涵盖了数据管理、业务分析和系统架构设计等领域,协同推进以下工作:(1)拓展数据应用场景,评估数据资源建设需求;(2)按需建设数据资源;(3)建立并维护统一的数据模型、数据标准,监测和提升数据质量;(4)建设并逐步完善数据治理规范体系;(5)搭建数据治理的总体技术架构,按需建设相关平台工具;(6)数据治理成效评价。其中,智图研究中心的9名外聘技术人员专职负责平台工具研发以及提供数据技术支持。

现阶段的数据治理组织架构和职责分工显著提升了馆员对图书馆全局业务和数据资源的理解,有助于其从图书馆全局视角提出融合技术、业务和数据解决方案,并快速、无缝对接业务环节。

4 成效与思考

4.1 成效

2022至2024年期间,复旦大学图书馆依托数字基座—数据中心,按需采集了14个图书馆业务系统和4个校内部门的数据,制定315项数据标准以及361条映射规则,建设了个人、文献载体表现和单件、师生行为等业务对象数据,发布了109个面向具体场景的应用数据集,总数据量近1亿条。截至2024年9月底,数据中心通过API接口为11个业务系统提供了累计约129万次服务(数据和应用场景见表1),同时支持以数据集下载的方式满足业务统计和分析的需求。

数据治理的具体成效体现在:(1)数据应用开发实现敏捷与高效。由于减少了从多系统采集和处理数据的工作,新的基于数据的应用系统(例如表1中的第5、6、10序号)的架构设计得以简化,因此开发周期相较以往平均减少三分之一,简单应用在一周内即可上线;(2)一数多用,充分发挥数据价值。2023年底,经过需求征集和评估,数据中心向馆内开放了48个应用数据集。在半年时间内,这些数据



集共收到 83 次下载申请,其中 21 个数据集的复用率在 50%到 83%之间。特别是针对馆藏和师生行为的多维度历年统计数据,不仅快速响应了数据需求,而且在多种业务分析和决策支持场景中得到了充分利用。此外,这些数据集正作为本地知识库,直接为基于大语言模型的智能咨询和资源推荐应用提供支撑;(3)助力传统业务提质增效。通过自助和订阅式的数据服务,业务部门现在能以更及时、便捷和灵活的方式获取数据。例如,定期自动生成的 CASHL 数据集,为业务部门节省了约 70%的人工数据处理时间,显著提升了工作效率。此外,基于算法的数据质量控制不仅实现了数据的全量审校,而

且取代了约 50%的形式审校工作,较好地解决了一些可能对应用效果产生较大影响的质量问题,例如中文书目数据出版日期的一致性和标准化问题。这些问题占到了全部数据质量问题的 25.11%。经过源头治理后,将有效提升数据在检索发现和馆藏分析评价等业务场景中的支撑能力;(4)数据驱动的服务创新。数据治理的成果使图书馆能够深入到教学、科研管理场景提供创新服务,例如,基于学术成果数据,图书馆与全校 28 个部处、院系进行了 200 多项委托和合作,包括学科规划与发展、人才评价与发展、国际合作与交流、经费测算与分配、评奖遴选辅助等。

表 1 数字基座—数据中心支撑的业务系统

序号	系统名称	数据	应用场景
1	复旦大学学术典藏平台	全校学者个人及其成果数据	多维度展示本校学术成果,提供不同口径的定制统计服务
2	复旦大学学术成果认领系统	全校学者个人及其成果数据	学术成果认领、维护和完善
3	移动图书馆	各分馆实时在馆人数,学生选课及课程指定参考书数据,师生行为定期统计数据等	图书馆资源和空间的一站式服务
4	复旦大学教学参考资源服务系统	纸电馆藏数据	面向研究生提供课程指定教参书全文服务,面向教师提供教参书录入辅助服务
5	图书馆数据大屏发布系统	文献资源利用统计数据	图书馆场馆大屏信息发布
6	图书馆数据库导航系统	电子资源元数据	电子资源导航服务
7	图书馆捐赠信息管理系统	纸电馆藏数据	捐赠资源数据与馆藏数据的查重
8	AchivesSpace 非书资料管理系统	个人实体数据	通过个人实体数据唯一 ID,以责任方式等关系,关联个人与非书资料以及更多馆藏资源
9	印谱文献虚拟图书馆图像管理系统	个人实体数据	通过个人实体数据唯一 ID,以责任方式等关系,关联个人与印谱、印章以及更多馆藏资源
10	图书馆读者积分管理系统	11 种师生行为定期统计数据	按既定规则定期计算用户的积分
11	图书馆机器人盘点系统	指定阅览室馆藏数据	比对盘点数据与馆藏数据

4.2 思考

回顾复旦大学图书馆近年来在数据治理方面的探索和实践,我们认为高校图书馆在实施数据治理过程中可能面临投入产出比、技术能力、治理效益和可持续机制方面的挑战,因此提出以下应对建议。

4.2.1 采用全局规划和局部精准治理结合的实施策略,提升数据治理投入产出比

数据治理是一项持续性的工作,除了软硬件投

入,还包括管理协调、流程优化、馆员培训等多重任务,其价值通常难以直接体现,成效也并非立竿见影。因此,对于经费有限的高校图书馆,采用全局规划和局部精准治理相结合的实施策略将有助于提升数据治理的投入产出比。

首先,将学校的发展战略和图书馆的转型需求相结合,明确数据治理的总体目标和主要任务。其次,全面盘点现有数据资源,对照治理目标,识别存



在的问题,并明确数据治理的优先领域、具体目标和业务价值评价指标。然后,聚焦关键数据,从业务痛点和师生急需的服务入手,并根据实际条件选择适合的治理方法,无论是轻量级的灵活的策略,还是全面系统的方法,都应确保数据治理成果能够快速转化为业务价值,以增强各部门对数据治理的认同和参与度,为进一步实施工作奠定基础。

4.2.2 强化多方合作,提升数据治理技术能力

高校图书馆需要应用专业的数据治理技术、工具和方法,才能有效推进数据治理的快速落地。复旦大学图书馆近年来通过校内外多渠道的合作,逐步增强了数据治理的技术能力。一方面,图书馆与学校数据中心合作确定个人及学术成果的数据定义、格式和交换协议,以促进校内跨部门的数据共享;另一方面,与行业企业建立长期合作关系,研发了与图书馆现有的IT架构、业务流程和基础设施深度集成的数据治理平台工具,确保技术解决方案能灵活应对业务需求和数据环境的不断变化。

目前,数据治理理论与实践主要针对结构化数据,而高校图书馆的学术资源通常以非结构化和多模态形式存在,因此,高校图书馆应加强与院系人工智能研究团队的合作,积极引入前沿技术以提升数据治理的效率和智能化水平。

4.2.3 增强用户参与和数据重用,提高数据治理对教学科研的支持效果

高校图书馆的数据治理工作还需进一步研究师生、科研人员以及机器等不同群体对于数据的具体需求,从而支持多学科和专业的用户以创新和创造性的方法来重用图书馆数据^[27]。例如,为研究人员提供详尽的数据细节以支持其深入探索研究问题;为机器学习算法提供易于处理的数据,如文本语料库、关系表、语义图、增强元数据等^[28]。为满足这些多样化的需求,图书馆需与研究人员、数据用户等密切合作,不断拓展和优化数据模型、数据标准、数据处理流程、数据服务能力、标准规范和平台工具等,增强其可访问性、可发现性、互操作性以及可计算性,为教学科研提供可信、可用、高质量的学术数据资源,进而持续提高数据治理的成效。

4.2.4 建立数据治理长效运营机制,充分发挥治理体系的效能

数据治理需要一个稳定且与时俱进的运营机制作为支撑,以确保其有效性和持续性。

首先,建立馆员数据素养提升机制至关重要。在数据治理起步阶段,复旦大学图书馆通过跨部门协作机制,重点培养“业务驱动的数据治理规划”和“数据驱动的业务优化和服务创新”能力,并逐步建立“数据驱动的工作模式”,为转型发展期的图书馆注入强劲的变革力量。在持续运营阶段,建议由专门的团队负责体系建设、能力构建、数据运营和数据文化推广等工作,面向全馆业务部门提供数据管理和应用的专业服务。

其次,需要构建组织、管理和制度等体系的协同改进机制。建议以业务痛点问题为突破口,将数据治理形成的流程、规范、平台工具等融入到业务环节,建立专题管理体系并落地相关措施。例如,复旦大学图书馆针对编目审校工作长期存在的低效问题,将数据质量自动监测和评价整合到业务流程中,减少人工审校并投入资源,加强源头治理,有效提升了工作效率和数据质量。

最后,建立以业务目标为导向的成效评价机制。通过定期评估一系列具体、可达且可量化的关键指标,如数据质量、数据驱动决策比例、数据复用率、数据自助使用率以及师生满意度等,识别并优化需要改进的领域。

5 结语

数据治理作为数据的核心管理手段和管理范式,是释放数据要素价值的基础和前提,对实现数据驱动式的战略发展至关重要。在教育数字化转型加速推进之际,复旦大学图书馆基于高校图书馆数据治理框架,聚焦多源异构数据的汇聚与融合,通过构建治理流程、平台工具、规范体系和组织架构,逐步建成标准统一、高质量的数据资源体系,为数据驱动的业务转型和服务创新提供了支撑。下一步,复旦大学图书馆将根据业务发展、技术进步和环境变化,不断优化和调整数据治理策略。希望复旦大学图书馆的实践探索可以为同行提供借鉴和参考,共同推进图书馆行业在数据治理领域的深入实践。

今天,以大模型为代表的生成式人工智能技术,已经对高校图书馆数据治理提出了新的挑战和需求。作为现有数据治理体系的重要补充,面向人工智能的数据治理(Data Governance for Artificial Intelligence,DG4AI)正受到业界的广泛关注。因此,



高校图书馆在积极拥抱大模型的同时,也应开展面向人工智能的数据治理研究和实践。这不仅可以提升大模型应用的准确性、可靠性和安全性,还能为图书馆的数据治理工作引入新的视角和方法。

参考文献

- 1 Grajek S. Higher education in 2023 and beyond: grand strategies for post-pandemic grand challenges [EB/OL]. [2024-07-15]. <https://events.educause.edu/-/media/files/events/webinar/2023/web2304/slides.pdf>.
- 2 Barber T, Barnett M, Groenhout R, et al. You can't have digital transformation without data governance [EB/OL]. [2024-07-15]. <https://er.educause.edu/articles/2024/6/you-cant-have-digital-transformation-without-data-governance>.
- 3 陈建龙,邵燕,张慧丽,等. 大学图书馆现代化指南针报告[J]. 大学图书馆学报,2022,40(1):22-33.
- 4 陈建龙,邵燕,张慧丽,等. 大学图书馆现代化的前沿课题和时代命题——《大学图书馆现代化指南针报告》解读[J]. 中国图书馆学报,2022,48(1):17-28.
- 5 安小米,许济沧,王丽丽,等. 国际标准中的数据治理:概念、视角及其标准化协同路径[J]. 中国图书馆学报,2021(9):59-79.
- 6 包冬梅,范颖捷,李鸣. 高校图书馆数据治理及其框架[J]. 图书情报工作,2015,59(18):134-141.
- 7 肖洁琼,奉国和. 国内外数据治理模型对比分析[J]. 文献与数据学报,2020,2(2):14-25.
- 8 吴锦池,余维杰. 图书馆数据治理成熟度评价体系构建[J]. 情报科学,2021,39(1):65-71.
- 9 王静,王鹏. 基于数字孪生的高校智慧图书馆数据治理模型研究[J]. 图书馆学研究,2023(11):41-50.
- 10 司莉,曾粤亮. 机构科研数据知识库联盟数据治理框架构建[J]. 图书馆论坛,2018(8):61-67.
- 11 顾立平. 数据治理——图书馆事业的发展机遇[J]. 中国图书馆学报,2016,42(5):40-56.
- 12 刘洁璇. 高校图书馆知识管理中的数据治理[J]. 情报科学,2018,36(1):108-112,129.
- 13 只莹莹. 中台在国家图书馆的应用思考[J]. 图书馆论坛,2021,41(10):76-82.
- 14 李琳,宛玲. 智慧校园背景下高校图书馆协同参与数据治理概念框架构建[J]. 大学图书馆学报,2023,41(3):79-87.
- 15 杨新涯,文佩丹,卓应忠. 智慧图书馆的全数据体系研究[J]. 图书情报工作,2023,67(13):29-35.
- 16 陈飞,黄文彬. 数字化转型背景下大学图书馆数据治理流程研究[J/OL]. 图书馆建设. <https://link.cnki.net/urlid/23.1331.g2.20240715.1847.004>
- 17 卢小宾,洪先锋,蒋玲. 智慧图书馆数据标准体系研究[J]. 图书

- 情报知识,2021,38(4):50-61.
- 18 刘倩倩,刘圣婴,刘炜. 图书情报领域大模型的应用模式和数据治理[J]. 图书馆杂志,2023,42(12):22-35.
- 19 禹作霖. 数字化转型背景下高校图书馆数据能力评价指标体系构建[D]. 上海:复旦大学,2023.
- 20 国家市场监督管理总局,中国国家标准化管理委员会. GB/T 34960.5-2018 信息技术服务 治理 第5部分:数据治理规范[S]. 北京:中国标准出版社,2018.
- 21 许鑫,兰昕蕾,邓璐芴. 数字孪生视阈下智慧图书馆业务融合研究[J]. 大学图书馆学报,2022,40(2):59-66,35.
- 22 田颖,杨滔,党安荣. 基于场景迭代的数字孪生城市构建逻辑[J]. 上海城市规划,2023(5):24-30.
- 23 IFLA. IFLA library reference model: a conceptual model for bibliographic information[EB/OL]. [2024-08-12]. <https://repository.ifla.org/handle/123456789/42>.
- 24 Fürber C. Data quality management with semantic technologies [M/OL]. Wiesbaden: Springer Gabler, 2016 [2024-08-15]. <https://doi.org/10.1007/978-3-658-12225-6>.
- 25 吴砥,饶景阳,吴晨. 教育大数据标准体系研究[J]. 开放教育研究,2020,26(2):75-82.
- 26 杨琳,王伟,诸纪,等. 面向数据共享的教育数据标准体系研究与建设实践[J]. 大数据,2020,6(6):3-13.
- 27 Candela G, Escobar P, Sáez M D, et al. A shape expression approach for assessing the quality of linked open data in libraries [J]. Semantic Web, 2023,14(2):1-21.
- 28 Padilla T, Allen L, Frost H, et al. Final report—always already computational: collections as data[EB/OL]. [2024-08-16]. <https://doi.org/10.5281/zenodo.3152935>.

作者贡献说明:

薛崧:选题和论文框架构思,论文撰写
张计龙:选题和论文框架构思,修改定稿
苗青:参与数据标准、数据模型和数据质量管理的研究和实践
任磊、赵敏、毛梦丹:参与数据治理平台工具研究和设计
张耀:参与数据治理平台工具研发

作者单位:薛崧、张计龙、毛梦丹,复旦大学图书馆、复旦—阿法迪共建智慧图书馆学研究中心,上海,200433
苗青、任磊、赵敏,复旦大学图书馆,上海,200433
张耀,复旦—阿法迪共建智慧图书馆学研究中心,上海,200433

收稿日期:2024年10月22日

修回日期:2025年1月15日

(责任编辑:支娟)



Construction of Data Governance Framework in Academic Libraries

— Practices and Reflections of Fudan University Library

XUE Song ZHANG Jilong MIAO Qing REN Lei MAO Mengdan ZHAO Min ZHANG Yao

Abstract: Under the background of the digital transformation in education and the paradigm shift in scientific research driven by artificial intelligence, data governance has become an inevitable choice for academic libraries to enhance their data management capabilities and cope with the transformation and development of modernization. Currently, research on data governance in the field of library and information science mainly conducted at the macro and theoretical levels. Few libraries consciously implement it as a systematic practice, and there is a lack of systematic guidance on data governance methods. This paper constructs a data governance framework for academic libraries based on the overall framework of GB/T 34960.5—2018. It references the key components of the general data governance framework and relevant national standards, as well as the development strategies and the challenges faced by academic libraries in current data management. The proposed framework consists of six key elements: strategic guidance, governance content, governance process, platform tools, governance norms and organizational structure. Taking the practices of Fudan University Library as an example, this paper explores key measures, challenges and countermeasures of data governance. Specific measures include: developing a data strategy to define goals, tasks, and implementation strategies; building a unified data model and implementing layered, transparent and differentiated processing to integrate multi-source heterogeneous data and create standardized, high-quality data; expanding application scenarios to drive data resource construction based on demands; improving data, models, tools, and processes through usage feedback and quality assessments to form a closed-loop governance process; constructing a comprehensive data governance platform for integration, management, and application to enhance data capabilities; gradually establishing normative documents covering management, data, technology, security, and tools to ensure that data governance is well-documented, feasible and controllable; forming a data governance structure with cross-departmental collaboration and dynamic role adjustments to promote library-wide data initiatives. Practice has attested that Fudan University Library's data governance has achieved remarkable results in aspects such as the agility of data application development, the full utilization of data, the enhancement of quality and efficiency of traditional operations, and data-driven service innovation. To address the challenges in data governance process, such as cost-effectiveness, technical capability, governance benefits, and sustainability, this paper puts forward recommendations. Key suggestions include adopting a strategy that combines holistic planning with localized precision governance to enhance cost-effectiveness; reinforcing multi-party cooperation with the university's informatization office, artificial intelligence research teams, and industrial enterprises to improve the technical capabilities of data governance; strengthening collaboration with researchers and data users to enhance support for teaching and research; establishing a long-term operational mechanism for data governance, including librarian competency enhancement, management coordination and improvement, and performance evaluation mechanisms to maximize the efficiency of data governance framework.

Keywords: Academic Library; Data Governance; Data Standardization; Data Governance Framework