

治理绩效

高校数据管理能力模型的分析与实施

陈旭

(中国政法大学发展规划与学科建设处, 北京 102249)

摘要: 随着大数据的蓬勃发展,基于数据管理能力的需求已逐渐渗透到各个领域。立足于对高校数据管理实际工作的深入思考,从全面优化高校数据管理的角度出发,首先分析并确认高校数据管理中3个核心问题;接着参考数据管理能力成熟度评估模型,结合高校数据管理能力现状,着重选取其中3个能力域逐项解决核心问题;最后,从上述能力域聚焦6个能力项,并规划高校数据治理的实施路径。

关键词: 高校数据; 数据管理; 能力模型; DCMM

中图分类号: G647; TP3-05 **文献标志码:** A **文章编号:** 1671-1807(2025)12-0326-06

教育数字化和信息化逐步推进,数据管理在高校发展中的地位愈发举足轻重。与此同时,高校不断丰富和增长的数据体系结构,与现有保守落后的数据管理能力和环境之间的矛盾也更加突出。由于当前数据管理缺乏统一规范和标准,导致信息障碍,产生数据孤岛,难以满足高校数据共享、优化资源配置和趋势分析等期望需求,阻碍高校管理和决策由粗放式向精细化发展。如何对教育领域繁杂的数据进行管理,已然成为当今诸多高校面临的课题。一套体系完善、技术成熟的数据管理能力模型,必将提升高校的数据资产能力,为优化资源配置和预测发展趋势提供决策分析,推进高校“双一流”建设和各项教育事业发展。

1 高校数据管理研究现状

教育部相继出台了《教育信息化 2.0 行动计划》^[1]《2020 年教育信息化和网络安全工作要点》^[2]等一系列国家政策文件,明确提出要全面利用大数据等信息化技术手段,不断提升高校教育数据管理能力。

从理论上,纵观学界对于数据管理能力的研究,多集中于政务^[3-4]、企业^[5-7]与医疗范畴,其中尤以教育领域的研究成果较为欠缺,且相关研究或大多为治理思路和价值的探讨,或者局限于图书馆^[8-9]、科研部门^[10-11]与财务部门^[12]等局部数据治理。上述研究较为零散片面,未能从根本上解决高校面临的数据困境。

从实践中,高校数据治理亦面临诸多挑战。组织体系上缺乏与多元数据治理主体相适应的统筹管理,流程机制不明确、管理规则不清晰,导致数据治理工作难以高效推进。此外,校内数据难以有效融合,校外数据未能及时共享,影响了数据价值的深度挖掘。

有鉴于此,本文立足于校级数据管理顶层设计,对标国家标准数据管理能力成熟度模型,并结合工作中的实践经验,探讨高等教育数据管理实施路径,以形成对该领域研究成果的补充。

2 高校数据治理的核心问题

众所周知,信息化时代数据通常具有 5V 特征,即大容量(volume)、高速度(velocity)、多样性(variety)、高价值(value)与真实性(veracity)。相较政府机构通过数据治理优化公共服务或辅助政策规划、大型企业通过用户画像,分析制定精准营销策略,实现个性化推荐和信息优化,建立用户满意度正循环和医疗系统利用临床数据对比及实时统计分析提高诊疗效率等成熟数据领域而言,当前高校数据治理成熟度尚有不及。

与前者相比,从数据数量而言,高校机构数据范围及数据等规模较小,通常维持在 TB(Terabyte)级别,达不到 PB(Petabyte)甚至 EB(Exabyte)级别。从处理速度而言,高校缺乏统一采集提取各部门、各学院及研究所数据的工具及方法,难以高效开展数据治理工作。从数据种类而言,受限于固有

收稿日期: 2025-01-20

作者简介: 陈旭(1984—),女,湖南长沙人,硕士,统计师,研究方向为统计、大数据和信息化。

数据处理方式,对数据表等结构化数据分析较多,文本、图片、音频与视频等半结构化数据和非结构化数据价值有待挖掘,因此数据分析结构单薄。从数据价值而言,缺少全校性数据标准整体规划,数据难以融合。从数据真实而言,缺乏数据的全局视图,无法进行审查和溯源。上述问题产生的原因主要归结于以下三个方面。

2.1 数据归口不清晰

随着办学规模不断扩大及管理职能逐渐丰富,高校组织机构也随之扩充增长,导致存在多个校内部门对同一数据都有操作权限,但具体归口责任未有明确划分。上述情况不仅可能导致数据管理的混乱,还可能引发数据不一致等质量问题。一方面,由于高校组织结构复杂多样,日常业务交流高频繁杂,校内各部门之间职责边界或管理范围有重复交叉,导致在校级层面执行数据统计时,对数据管辖归属不明确,存在不同部门对同一数据进行重复管理。例如高校资产部门和基建部门同时管理高校各类建筑物,在《高等教育事业统计》《投资项目申请表》相关报表及台账统计中,对于某项新增建筑的投资金额、竣工时间和占地面积等数据,是由资产部门或者是基建部门进行归口统计。例如,人事部门和组织部门同时掌握高校教职工数据,当年新入职教职工党员数量,是由人事部门或者组织部门归口统计。另一方面,各部门业务内涵差别导致对于同一个数据,不同部门之间存在不同的数据结构与统计口径,导致不同部门对同一个数据产生不同的数据描述。以当年入学新生为例,高校招生部门统计数据通常从招生录检表中摘取,即当年录取报到的学生;而各学院统计数据通常中学籍管理系统中摘取,即当时实际在读的学生。由于二者统计时间和范围等口径不一致,其中部分退学或是补录的学生是否纳入数据范围存在争议。此外,当两个以上部门同时对一项数据有操作权限时,一个部门修改数据信息,更新后的内容未必同步至其他相关部门。例如,每年毕业季临近高校暑假,各学院统计本院学生就业情况时,通常不会与学生就业部门数据实时同步,由此造成校内数据存在误差。基于以上原因,高校数据在管理时存在数据归口不清晰,产生数据冗余或缺失,导致数据不一致,影响数据质量。

2.2 数据结构不统一

由于信息化技术持续更新迭代,高校在信息化建设并不是一蹴而就,而是一个循序渐进的长期过程。高校在建设初期,往往缺乏对数据结构和标准

的统一规划,给后续数据集成工作带来制约。一方面,高校日常数据平台通常包括教务、人事、财务、资产、科研等业务活动。虽然大部分高校已将以上平台统一集成为高校统一办公平台。然而,随着信息技术深入发展,高校业务活动产生的数据种类日益丰富,数据结构千差万别。高校固有办公平台数据结构相对固定,可扩展性较差,难以满足不断增长的数据接入需求,形成校内“数据孤岛”。根据调查,约有六至七成高校建立了高校统一办公平台,但已建成平台中仍有八成以上无法完全兼容新增业务系统或是用户数据。以上平台或因体系结构庞大落后,或因代码逻辑复杂难以修改,后续升级改造难度也较大,以致用户在进行日常数据管理时,不仅需要在多个系统中来回切换,并且只能通过人工手段同步补全及更新数据,特别是临近学年开始和结束的业务高峰期,使得本就忙碌的工作更加繁冗。另一方面,高校在日常办公中,高频使用上级主管部门,如区市教委、教育部、财政局、统计局和发改委等政府机构开发的信息系统,以上系统各自为政,且与高校办公平台之间数据结构各异。高校在完成上述系统填报任务中,可能花费大量人力精力采集了不少有价值的信息,但囿于数据结构之间的差异,难以有效整合和利用,未能集成接入高校数据中心,使得大量有效数据信息仅仅提供给上级主管部门使用,而无法与校内已有数据良好融合,导致大量数据被孤立,并未发挥其应有的价值,形成二次或多次利用,造成数据资源浪费。

2.3 数据应用不全面

相较于政府、企业与医疗等数据成熟应用行业,由于缺乏管理规范和技术支撑不足等因素,导致高校数据价值挖掘度较低,数据应用不全面。依照其来源、结构及存储方式,数据通常被划分为结构化数据、半结构化数据和非结构化数据。一方面,高校结构化数据层次单一。高校结构化数据来源多为《教育事业综合统计》《全国普通高校科技统计》《“双一流”建设监测数据》《高等教育质量监测数据统计》等校级数据报表,以上校级报表虽然数据范围广,涵盖高校学生、师资、学科、科研、资产等整体指标,但数据颗粒度较大,大部分指标无法继续下分至各学院。又出于数据安全及保密规定,限于缺乏技术途径做数据脱敏,校际数据流通受限,各校无法掌握兄弟院校同级数据或区市级教育部门上级数据。只能就自身发展情况做简单增速比较,无法与同级或上层级高等教育发展情况进行横

向对比。另一方面,半结构及非结构化数据未能有效规范化。高校半结构及非结构化数据主要来源于教学过程中的课程记录;行政办公过程中的相关文档、图片和视频以及校园生活中的行程轨迹与监控影像等。以上数据有增长速度快、单体数量大、结构差异大等数据特征。虽然大数据与人工智能等前沿知识已在高等教育教学,特别是课程理论方向,实现了高度普及。然而大部分高校在此项管理和应用上的实践仍有滞后。因此,高校对于分散存放在不同平台和系统中的半结构化和非结构化数据,由于其数据结构不规则或不完整,没有预定义的数据模型,缺乏成熟统一的管理标准和规范,造成数据资源闲置。而且,传统的数据库工具无法直接处理半结构化和非结构化数据,大部分高校,特别是非理工类高校,缺乏先进的数据分析实践技术,不能对半结构化与非结构化数据进行大量、高效、实时处理,以致以上数据分析与决策等实际应用延迟或是无效,使得数据潜在价值和深层信息难以挖掘,进而影响数据应用的广度和深度。

3 高校数据管理的参考模型

数据管理能力成熟度评估模型^[13](data management capability maturity assessment model,DCMM)是由全国信标委大数据标准工作组研发,并于2018年3月15日正式发布,是我国在数据管理领域首个正式发布的国家标准,旨在利用先进的数据管理理念和方法,建立和评价自身数据管理能力,持续完善数据管理组织、程序和制度。DCMM是一个整合标准规范、管理方法论、评估模型等多方面内容的综合框架,它将组织内部数据能力划分为数据战略、数据治理、数据架构、数据应用、数据安全、数据质量、数据标准、数据生存周期8个数据能力管理域,每个能力域又包含3或4个能力项。

目前,DCMM在政府^[14]、企业^[15]及医疗^[16-17]等领域均有成熟应用,并通过案例实践证明其对数据管理具有有效指引。同样,DCMM针对当前高校数据归口不清晰、数据结构不统一、数据应用不全面的问题给出解决问题的过程描述与目标标准等实施路径。高校在数据管理过程可以因地制宜,针对薄弱环节查漏补缺,建议高校在此过程中应有所侧重,着重数据治理、数据架构与数据应用3个数据能力管理域以解决数据管理短板。

4 高校数据治理的实施路径

4.1 通过数据治理能力域解决数据归口不一

数据治理能力域是数据管理框架的核心职能,

是对数据资产管理执行权利和控制的集合,此能力域通过建立健全数据管理的组织、流程、架构等多个方面,实现数据来源定岗定责与归口管理。此能力域包括数据治理组织、数据制度建设、数据治理沟通3个能力项。相对于政府部门、大型企业与医疗机构,高校优势为人员数量相对较少、组织结构相对简单,沟通成本较低及流转程序便捷,劣势在于缺乏统一完善的数据管理机构与成熟可行的数据管理制度。因此在此能力域中,薄弱环节主要存在于前两个能力项。

数据治理组织能力项包括组织架构、岗位设置、团队建设、数据责任等内容,是各项数据职能工作开展的基础。高校可以通过以下路径完成此能力项。第一,建立数据管理组织,高校可以成立以统计或数据部门分管校领导为首,相关统计或数据责任部门领导为成员的数据管理领导小组,下设数据管理办公室。第二,明确数据管理办公室的业务职责与岗位编制,该办公室可以负责统筹高校常规综合性报表报送以及数据服务工作。第三,强化数据管理团队建设,通过设立校级数据联络员制度,在每个数据责任部门选聘一名数据联络员,给予一定资金与政策支持,并依靠上级部门、同级高校交流培训强化业务能力。第四,数据归口管理,根据高校数据需求和上级部门标准逐项分解高校数据,并按照部门业务层层落实数据管理职责,确认数据联络员负责数据范围、口径及更新频率,并定期考核评审。

数据制度建设能力项可以保障数据管理和数据应用各项功能的规范化运行,是数据管理和数据应用各项工作有序开展保障。高校在此能力项实施过程中,可以参考《中华人民共和国统计法实施条例》(2017年)、《教育统计管理规定》(2018年)等上位法规条例,并根据本校的机构设置,重新确定各级部门在数据管理中的任务、职责、报送及发布流程等具体要求,可就责任、签审、保密、申请、共享等若干方面建立制度约束。高校在此能力项中,应根据数据职能的层次和授权决策次序,将数据制度框架细分为政策、办法、细则三个层次^[13]。其中,数据政策是针对数据管理和应用的目的,明确其组织和范围,是数据制度能力建设的顶层设计,应由数据管理办公室制定,数据管理领导小组审议通过。数据管理方法是完善高校数据管理和数据应用各领域,如教学和科研活动开展而制定的相关规则和流程,是数据制度能力建设

的中坚支撑,应由各二级数据责任部门制定,数据管理领导小组审议通过。数据管理细则是为确保各数据方法执行落实的相关文件,是数据制度能力建设的基础保障,应由各二级数据责任部门制定,数据管理办公室审议备案。高校数据管理制度可首先在高校数据管理领导小组及相关部门内广泛征求意见并发布试行。然后通过参考兄弟院校先进经验和数据行业领先标准持续优化,实行数据制度动态更新机制。

在完善数据治理能力域建设过程中,首先设立信息统计专职部门及岗位,明确该专职部门的工作内容与职责范围;其次,根据相应统计或数据任务,实现单项任务数据归口管理;再有,依照上位法律及实施条例,制定校内信息统计工作管理规定。与进行此项能力与建设之前相比,数据管理有岗归责、有章可循,工作流程也更加便捷高效。然而,由于组织架构的顶层建设欠缺,例如,未成立数据管理领导小组统一部署协调,因此少数使用频率较低数据的归口责任部门仍未明晰。

4.2 通过数据架构能力域解决接口不统一

数据架构能力域用于定义数据需求,指导对数据资产的整合与控制,使数据投资与业务战略相匹配的一套整体构件规范,此能力域可通过维护校级数据模型一致性,解决数据接口不统一等问题。此能力域包括数据模型、数据分布、数据集成与共享和元数据管理4个能力项。当前高校数据分布关系较为明确,元数据管理需求较低,困境在于来源多,没有统一数据标准模型,制约了数据资产价值,难以集成共享,无法实现组织内部数据高效互联互通。因此在此能力域中,应侧重数据模型和数据集成与共享能力项。

数据模型能力项是使用结构化的语言将收集到的组织业务经营、管理和决策中使用的数据需求进行综合分析,按照模型设计规范将需求重新组织。高校可以通过以下路径完成此能力项。第一,获取和分析数据需求。高校应该由数据管理办公室根据常规数据采集任务和临时数据采集业务,汇总、梳理,并确认高校数据需求,包括满足教学和科研的高校内部业务需求,以及上级主管部门和兄弟院校共享的外部交流需求。第二,制定模型规范并开发数据模型。根据数据需求和业务实际,定义校级数据结构和数据字典,明确数据模型的命名规范及管理方法,建立校级数据的概念、逻辑与物理模型与映射关系,同时指导二级业务部门建立部门数

据模型。第三,应用并检查数据模型,检查数据模型与数据需求的一致性,确保各级数据模型可以有效满足已确认的数据需求。第四,维护并优化本校数据模型,即根据高校内外部数据需求变化,调整数据模型,并建立追踪与审查机制。

数据集成与共享能力项是建立起组织内各应用系统、各部门之间的集成共享机制,通过组织内部数据集成共享相关制度、标准、技术等方面的管理,促进组织内部数据的互联互通。高校可以建立数据集成共享制度、流程和环境。高校可以通过以下路径完成此能力项。第一,明确数据共享的原则和范围,例如校级师生等共有数据以及论文等外网公开数据。第二,建立共享流程,例如数据申请由实名制职工提出申请并列明用途,由其所在处级分管领导签名确认,后交由数据归口管理单位处级领导审批交数据联络员提供并由数据管理办公室存档。第三,建立数据共享环境,以上环节可通过集成于高校统一办公平台实现。第四,健全数据同步机制,数据归口部门变更数据后,主动将相关数据信息同步至相关业务部门,实现数据更新零延迟,消除高校数据不一致性。

在完善数据架构能力域建设过程中,首先根据常规数据采集任务和部分内部业务,分析确认此项需求;其次,建立健全校内信息共享制度及工作流程。与进行此项能力与建设之前相比,数据的需求管理更加完善,共享途径更加顺畅,然而,由于未能建立校内数据模型和同步机制,导致数据资产价值降低和校内数据不一致。

4.3 通过数据应用能力域解决应用不全面

数据应用能力域是指通过对高校校级数据进行统一的管理、加工及应用,此能力域通过支持数据分析和数据开放共享,挖掘数据价值,从而避免数据应用不全面,提升数据在高校各项事业发展的支撑辅助作用。此能力域包括数据分析、数据开放共享、数据服务3个能力项。高校数据分析能力普遍较弱、受限于不同校内报表体系与信息系产生的数据孤岛,开放共享能力有待提升。另外,高校数据应用偏向于内部决策参考而非外部服务。此能力项应着重于数据分析和数据开放共享能力项。

数据分析能力项通过数据分析或挖掘建模,对高校各项决策活动提供数据支撑。高校可以通过以下路径完成此能力项。第一,深化结构性数据分析,可对《教育事业综合统计》《全国普通高校科技

统计》等常规报表进行统计分析,分析各类报表数据关系,厘清同类统计项之间的数学联系,建造校级结构化数据分析体系。第二,布局半结构化和非结构化数据分析,通过文本挖掘、自然语言处理、图像处理和行为识别等大数据与人工智能技术,对教学计划、科研报告、实验数据和课程影音等教学、科研数据进行处理,提取海量数据中价值信息,建立校级分析决策神经网络。第三,融合结构化数据、半结构化和非结构化数据,整合数据资产,并在校级层面建设统一数据平台,实现对以上数据有机管理,支持跨部门及部门内部常规数据分析和数据开发。第四,对于常规分析建立统一数据分析应用的管理办法,对于非常规且跨多部门的分析需求,可采用点对点或星型数据联通方式处理。

数据开放共享能力项按照统一的管理策略对高校内部数据进行有选择的对外开放,同时按照相关的管理策略引入外部数据供高校内部应用。高校可以通过以下路径完成此能力项。第一,可通过数据管理办公室获取校内常规数据需求,并审查其合理性及可行性;第二,将以上需求分工落实至各数据归口管理部门,并结合《高等学校信息公开办法》(2010年)形成校级开放数据目录;第三,数据管理办公室制定校内外权限开放共享策略,对数据实现分级开放、统一管理,规范口径及更新频率;第四,高校可建立外部数据共享清单,及时获取上级部门和兄弟单位对标数据,通过内外部发展比对,更全面地了解各项事业发展水平。

在完善数据应用能力域建设过程中,首先,通过综合若干常规性报表,形成校级结构化数据体系;其次,建设校级统一数据平台,实现了校内数据管理;再有,制定了学校信息公开目录,并依照此目录定期对外更新数据。与进行此项能力与建设之前相比,拓展了结构化数据分析的维度,为后续数据管理可持续建设奠定基础。然而,由于未能深入挖掘半结构化与非结构化数据信息,未能深度整合校内资产,导致全局数据分析范围受限,数据应用能力有待提升。

5 结论

高校数据管理能力成熟度模型基于DCMM,并根据高等教育组织机构、技术能力和数据现状有机搭建。依托此模型参考比照,旨在查找当前高校数据管理中的短板与缺陷,并依照实施路径,有效破局高校数据困境,向高校持续提升数据管理能力提供理论支撑与实践参考,从而为高校各项事业发展

提供更高效可信的数据辅助,推动高等教育数字化转型和智能化发展。

参考文献

- [1] 教育部. 教育部关于印发《教育信息化2.0行动计划》的通知[J]. 中华人民共和国教育部公报, 2018(4): 118-125.
- [2] 教育部. 《2020年教育信息化和网络安全工作要点》[J]. 中国信息安全, 2020(3): 17.
- [3] 陈文明, 程永红. 数字政府建设下地方政务数据应用的标准化对策研究[J]. 标准科学, 2023(11): 65-70.
- [4] 万方, 周西平. 基于DCMM的我国警务数据管理能力成熟度评估[J]. 云南警官学院学报, 2021(1): 90-96.
- [5] 王平. X公司产品数据治理体系的构建与应用研究[D]. 大连: 大连海事大学, 2019.
- [6] 王畅. 徐工集团数据管理能力成熟度评估及优化路径研究[D]. 徐州: 中国矿业大学, 2023.
- [7] 李唐, 李青, 陈楚霞. 数据管理能力对企业生产率的影响效应——来自中国企业-劳动力匹配调查的新发现[J]. 中国工业经济, 2020(6): 174-192.
- [8] 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.
- [9] 陈媛媛, 林安洁. 高校图书馆科研数据管理服务模式搭建和应用[J]. 情报理论与实践, 2023, 46(5): 99-106.
- [10] 崔涛, 李青. 比利时高校科研数据管理服务调查研究[J]. 情报杂志, 2023, 42(10): 185-193.
- [11] 开放科学背景下高校科研数据治理模型的构建与管理路径[J]. 中国管理信息化, 2024, 27(23): 109-112.
- [12] 张牧. 开放科学背景下高校科研数据治理模型的构建与管理路径[J]. 中国管理信息化, 2024, 27(23): 109-112.
- [13] 《数据管理能力成熟度评估模型》: GB/T 36073—2018[S]. 北京: 中国国家标准化管理委员会, 2018.
- [14] 马达, 严玲玉. 基于DCMM的政府数据治理路径探究[J]. 网络安全和信息化, 2023(8): 46-48.
- [15] 高智伟, 杨凯, 张智祥, 等. DCMM引领企业数字化转型路径研究[J]. 中国经济报告, 2022(5): 138-143.
- [16] 林琳. 基于DCMM框架的医疗数据治理体系研究[J]. 网络空间安全, 2024, 15(4): 361-366.
- [17] 谭坤, 袁博, 毛云鹏, 等. 基于DCMM的医院数据管理能力成熟度评估研究[J]. 卫生软科学, 2024, 38(5): 43-47.
- [18] 边香玉. 大数据时代高校信息管理与决策机制研究[J]. 无线互联科技, 2020, 17(23): 48-49.
- [19] 余鹏, 李艳. 大数据视域下高校数据治理方案研究[J]. 现代教育技术, 2018, 28(6): 60-66.
- [20] 汤贝贝. 大数据背景下高等教育治理转型: 机遇、挑战与应对策略[J]. 重庆高教研究, 2019, 7(2): 77-86.
- [21] 郑苑, 梁振辉. 教育信息化背景下高校数据治理研究[J]. 中国教育信息化, 2020(17): 50-54.
- [22] 张培, 夏海鹰. 教育领域数据治理的基本思路与实践路径[J]. 现代教育技术, 2020, 30(5): 19-25.

Analysis and Implementation of University Data Management Capability Model

CHEN Xu

(Development Planning and Discipline Construction Department, China University
of Political Science and Law, Beijing 102249, China)

Abstract: With the booming development of big data, the demand for data management capabilities has gradually penetrated into various fields. Based on a deep reflection on the practical work of data management in universities, starting from the perspective of comprehensively optimizing university data management, three core issues in university data management were firstly analyzed and confirmed. Then, referring to the maturity evaluation model of data management capabilities, combined with the current situation of university data management capabilities, it focused on selecting three capability domains to solve the core issues one by one. Finally, from the above capability domains, it focused on six capability items and plans the implementation path of university data governance.

Keywords: university data; data management; capability model; DCMM