

# 开放大学数字图书馆数据治理探析\*

付博 武兴华

(国家开放大学, 北京 100039)

**[摘要]** 文章将国家开放大学数字图书馆现有业务特点和发展重点作为研究内容, 在阐述数据治理概念及研究基本情况的基础上, 分析开放大学数字图书馆发展情况与存在问题, 根据其自身特点提出有针对性的数据治理框架, 意在提升开放大学数字图书馆数据资产的服务水平, 提高数据的利用率和数据价值。

**[关键词]** 开放大学; 数字图书馆; 数据治理; 数据质量

**[中图分类号]** G258.6

**[文献标识码]** A

**[文章编号]** 1008-7656 (2019) 01-0081-05

在信息化与数字化的技术驱动下, 数据科学时代拉开序幕, 新的研究范式应运而生。随着数据科学不断发展, 管理过程设定、业务流程决策以及数据处理方法也在时刻发生变化。异构海量数据的产生使传统的数据管理方式“捉襟见肘”, 亟需对数据进行科学治理以确保数据的高效利用与最大增值。

## 一、数据治理的概念及数据治理研究的基本情况

### (一) 数据治理的概念

目前, 国内外的研究机构和学者从不同视角对数据治理进行定义, 但还未形成统一共识。一些国外学者从法案中提出数据治理, 定义了一系列规则和政策<sup>[1]</sup>; 国际数据管理协会(Data Management Association)认为, 在数据管理的十大职能中, 数据治理最为核心, 在数据的管理和使用层面对数据进行规划、监督及控制, 用以支持顶层决策与战略实施<sup>[2]</sup>; 还有一些学者主张数据治理

应从数据管控的责任与技术等角度出发, 在人与技术的集中使用背景下, 组织中的数据作为一项资产得到合理使用<sup>[3-6]</sup>。美国全球数据管理协会(DAMA)认为, 数据治理是对数据资产管理行使权利和控制的活动的集合(DAMA, 2012)<sup>[7]</sup>。国际数据治理研究所(DGI The Data Governance Institute)认为, 数据治理是指针对信息相关过程的决策权和职责体系, 这些过程遵循“在什么时间和情况下、用什么方式、由谁、对哪些数据、采取哪些行动”的方法来执行<sup>[8]</sup>。在国内, 数据治理的定义还没有统一, 多数研究与数据的生命周期相关, 例如数据的采集、加工、存储、利用等, 或采用数据管护、数据管理等名词<sup>[9-14]</sup>, 还未提到数据治理的本质。数据治理源于IT治理, 其将数据作为一项有重要价值的资源, 不仅仅是对数据管理过程的计划执行, 更是对管理过程的指导、监督和评估。

### (二) 数据治理研究的基本情况

框架方面, 除DGI框架和DAMA框架<sup>[15]</sup>外,

\* [基金项目] 2017年国家开放大学专项课题“嵌入数字化学习空间的数字图书馆服务研究”(编号: G170011S)  
[收稿日期] 2018-11-05

国外也有学者提出了决策域模型和驱动力模型<sup>[16][17]</sup>。国内包冬梅等总结了数据治理框架中的职能及关系、工作区间、任务、组织结构、责任分工、成效的评估标准等,同时提出了高校图书馆的数据治理框架 CALib<sup>[18]</sup>;许晓东等设计了高等教育数据治理的分析框架<sup>[19]</sup>;李青、韩俊红设计了教育数据治理框架<sup>[20]</sup>。

在图情领域,顾立平从数据获取、数据共享、数据重用三方面证实了数据治理是图书馆事业发展重要的机遇<sup>[21]</sup>;刘洁璇阐述了数据治理在知识管理中的应用与价值,并讨论了发展前景<sup>[22]</sup>;任亚忠分析了大数据时代背景下,数据治理为图书馆带来的一系列转变<sup>[23]</sup>。

## 二、开放大学数字图书馆发展情况

开放大学采用多媒体数字化资源办学的特点,决定了信息化与数字化将是支撑其远程教育的核心与焦点。数字图书馆作为开放大学重要的组成部分,在现代信息技术的支撑下,采用网络化、数字化提供服务已是常态。

### (一) 业务特点——海量数据与大数据利用

开放大学的学生数量庞大,分布在不同的年龄层次,也遍布不同的城乡区域、机构岗位与社会阶层,知识背景与学习能力也各不相同,同时,学生学习规律、认知习惯、学习动机、学习环境方面也差异明显。因此,每个学生在使用数字图书馆时,存在不同的学习行为与学习路径,构成了海量读者行为数据的来源。与此同时,在大数据飞速发展的背景下,如何将这些行为数据加以价值最大化利用,这成为了目前的研究热点并值得深入挖掘。

### (二) 发展重点——数字资源深度组织与开放获取

数字图书馆是用户获取数字资源及相关服务的平台。依据传统的服务模式,数字资源通过数据库和网页形式向用户呈现,根据数字资源自身主题或学科导航进行资源揭示,由用户主动进行资源获取,该模式下的需求导向不足。同时,数字资源整合深度难以满足用户的个性化需求,挖掘深度较浅。开放大学强调优质资源的共建

共享,将免费开放与传播的公共学习资源纳入数字化服务中,契合开放大学的办学理念与服务宗旨,能够推动开放大学数字图书馆资源整合空间的无限延伸,开放获取将得到拓展和开辟。

## 三、开放大学数字图书馆数据治理框架

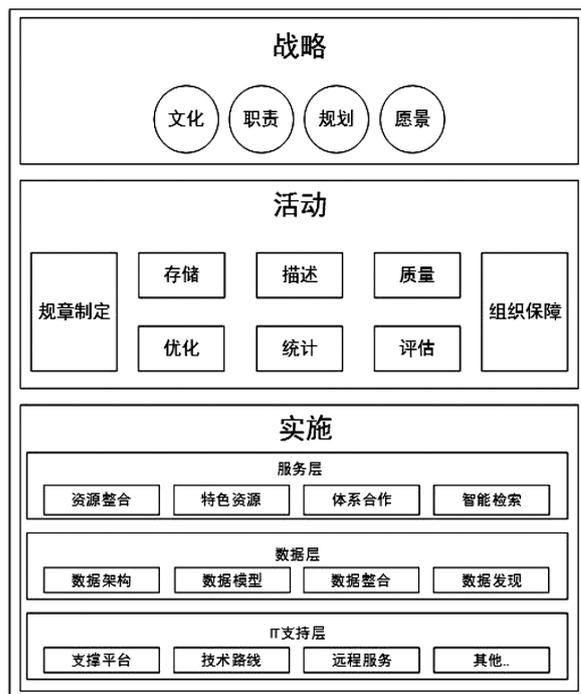


图1 开放大学数字图书馆数据治理框架图

开放大学数字图书馆数据治理是在开放大学图书馆发展战略和规划的指导下实施的,因此,其数据治理工作具有鲜明的行业特征,通用的数据治理框架并不完全适用。通过上述对开放大学数字图书馆的业务特点和发展重点进行分析,提出开放大学数字图书馆数据治理框架(见图1)。

从内容上来看,开放大学数据治理包括数据治理战略、数据治理活动体系、数据治理实施三个层面。在这三个层面中,包括数字图书馆的文化内涵与愿景、规章制度与组织保障,以及资源服务、数据整合、基础IT支持等多个方面。

### (一) 数据治理战略

数据治理战略处于框架的最顶层,对整个数字图书馆的发展具有最高指导意义。数字图书馆秉持“时时、处处、人人”的服务理念,为学生

终身学习打造优质平台,并逐步适应目前先进的技术与大数环境,与时俱进,使各类学习者不受空间时间限制,随时随地获取数字资源服务。同时,数据治理的战略意识也应当贯穿到数字图书馆建设的各个层面,以战略为指引,形成具体活动与举措的规划。

## (二) 数据治理活动体系

### 1. 规章制度与组织保障

数据治理覆盖面广,各项规划与举措,各类业务和流程,各种系统与平台,要求每项工作都必须从制度上加以规范与保障,将各类章程、办法、考核等贯穿到各项工作的始终。在规章制度的指导下,进行人力财力等的有效组织,保证数据治理工作的顺利进行。根据学校的愿景和目标,设定总方向,获得领导层的支持;进行跨部门的协调,执行数据治理具体计划;业务部门划定具体业务的负责工作组或工作人员,履行职责,解决问题。

### 2. 数据存储治理

对于数字图书馆来说,各类数据的存储安全是最不能忽视的一个问题。数据的存储是其持续利用的关键点,存储安全应当得到优先保证,并且需从以下几点着手并逐步制度化、常态化。

(1) 安全规划。安全规划应保证存储在数据库中的所有数据安全有效,在经过长时间存储之后仍能被访问和使用。具体措施如定时评估数据、监测技术环境、测试存储功能、对所部署的安全措施进行审计、注意数字资源的权限管理等,了解可能发生的安全事件。

(2) 日志分析。日志被认为是一类非常重要的数据,它显示出了数据的存储方式、存储路径等,基于日志的分析能够帮助系统管理员更好地了解数据访问使用情况,了解所出现的问题并能及时查看。

(3) 访问控制。为了防止数据被不正当访问,减少数据滥用与数据泄露发生的可能,需要对数据的访问进行权限设置,某些权限只能授予那些需要访问数据的人,以此来降低越权访问率与事故后的处理难度。

(4) 容灾备份。容灾备份实际上提供了一套复制数据的机制,将数据保存在另一个物理独立的存储容器中。容灾备份对所有通信和数据定期检查,以确保在数据库运行崩溃时,能够在最短时间内使资源得到安全、高效的恢复。

### 3. 数据描述治理

(1) 数据编码标准。建立统一的资源数据编码标准可以保证在全体数字资源的范围内,将数据进行一致性管理,为实现数据库的数据集成共享,管控协同,一体化运行做好良好的基础保障。

(2) 数据编码原则。结合数字资源管理需求,制定数据编码规则如下。

识别唯一性:统一资源数据编码要求具有唯一的识别标志。

特征一致性:保证资源数据编码的关键特征在不同应用、不同系统中的高度一致性。

数据稳定性:资源数据编码本身的属性不会随管理过程中的数据流转而被修改。

长期有效性:资源数据编码所代表的数据对象在存储中仍然存在,该组员数据编码就应该保持有效性。

(3) 元数据描述标准。元数据是具有对数据进行定位、描述、选择等多种功能的数据,具有连贯有效地描述、管理、编目网络资源的功能。通过对元数据的管理,对数据进行分类组织,实现无缝信息检索、内容调用和资源共,是数据治理的重要内容。在元数据管理中,可以制定三大类元数据项:必选项(必须有)、条件项(有则应用)和可选项(有用时)。

### 4. 数据质量治理

数据质量治理是提高和改善数字资源数据质量的一项活动,是对数据的每个阶段——采集、存储、管理、共享到应用中可能引起的各种数据质量问题进行管理,从而进行资源监测、问题识别与事故预警。质量治理是一个循环管理,其最终要达到的目标是提升数据质量,通过可靠的数据提升其使用价值,并最终获得管理效益。

数据的质量治理最主要包括完整性、准确性与有效性等三个方面。

(1) 数据的完整性。数据的完整性就是数据之“全”。即在数据生命周期开始时,把所需信息尽可能采集完整,这对于全面把握数据,对数据进行分析并发现其中规律及问题具有重要意义。随着数据表格、数据库的深度使用,能够对数据完整性进行完善。

(2) 数据的准确性。数据的准确性是指数据是否符合其真实的情况。对于数字资源来说,其数据的准确性尤为重要。因此,我们进行资源数据管理时,在整合、传输等各个环节,都有可能受到各个方面的影响,因此我们要尽可能将误差控制到最小,以保证数据的准确性。

(3) 数据的有效性。在保证数据的完整性与准确性的同时,还需要考虑数据的有效性。也就是说,我们在衡量数据时,过期的、不符合要求的数据是不产生价值的,是应该被剔除的。

#### 5. 统计、评估与优化

数字图书馆数据采集系统将用户的行为数据记录下来,并进行统计分析,直接服务于提升数字图书馆的建设水平和服务质量的活动。在加强统计分析与数据挖掘的同时,推动学科服务、知识服务等完善。建立可实行的、可测量的统计与评估指标体系,了解数据治理的现状和差距,对数据治理框架中的各项内容进行不断优化,制定改进路径,促进数据治理向成熟发展。

#### (三) 数据治理的实施

数据治理的实施指在战略规划及关键活动指导下,在IT支持层面、数据层面和服务层面所做的具体措施。

IT支持层面包括技术路线、支撑平台、远程服务等,在支撑整个数字图书馆系统运行的基础上,针对开放大学师生遍布全国的特点,提供远程登陆、联合认证等服务;数据层面包括数据模型、数据架构、数据整合、数据发现四个方面。数据架构是系统和软件架构层面的描述,主要是从系统设计和实现的视角来看数据流和信息流。数据模型包括概念数据模型、逻辑数据模型和物理数据模型等。同时,在标准化、开放性与集成化的标准环境下,整合数据库的各类资源、整合

纸本与数字资源、整合专业、小众的学科资源提供知识路径等,也正在成为数字图书馆系统发展的主流;服务层面包括构建基于开放网络的、适合开放大学特点的学术资源体系,进行资源整合,建立特色资源库,采用最适合的元数据方案,加强数字资源元数据编目,针对成人教育的特点,对资源进行更细粒度、更小单元、更深层面的组织和揭示。与此同时,在开放大学系统中,充分挖掘各分部发展需求,整合各地优势资源,为整个开放大学体系提供服务。

#### 四、结语

开放大学数字图书馆随着技术与数字资源的不断前进演变,越来越以数据为核心。随着国家“十三五”规划进入关键时期,数字图书馆的发展与建设也应该站在“十三五”规划的高度,对数据治理加以重视。应树立数据治理的理念,在新环境中进行转型,在总体战略的引领下,明确职责分工,加强组织保障,完善标准体系,提供高质量的数据服务,将数据作为宝贵资产加以利用,为开放大学的师生提供便捷、满意的服务。

#### [参考文献]

- [1] Donaldson A, Walker P. Information governance—A view from the NHS [J]. International Journal of Medical Informatics, 2004 (73).
- [2] 国际数据管理协会 [EB/OL]. (2015-05-31) [2018-10-12]. <https://dama.org>.
- [3] Fernandes L, O'Connor M. Data governance and data stewardship. Critical issues in the move toward EHRs and HIE [J]. Journal of AHIMA/American Health Information Management Association, 2009 (80).
- [4] Rosenbaum S. Data governance and stewardship: Designing data stewardship entities and advancing data access [J]. Health Services Research, 2010 (45).
- [5] Haider A. Asset lifecycle data governance framework [A]. Proceedings of the 7th World Congress on Engineering Asset Management (WCEAM2012)

- [C]. Switzerland:Springer International Publishing, 2015.
- [6] Malik P. Governing big data: Principles and practices [J]. IBM Journal of Research and Development, 2013 (3) .
- [7] DAMA International.The DAMA Guide to the Data Management Body of Knowledge [M]. 1st ed.New York: Technics Publications, 2009.
- [8] Data Governance Institute.The DGI Data Governance Framework[R/OL].(2015-08-20)[2018-10-10]. <http://www.datagovernance.com/the-dgi-framework/dgi-framework.pdf>.
- [9] 程莲娟.美国高校图书馆数据监护的实践及其启示 [J].图书馆杂志, 2012 (1) .
- [10] 钱鹏, 郑建明.高校科学数据组织与服务初探 [J].情报理论与实践, 2011 (2) .
- [11] 丁培.数据策展与图书馆 [J].图书馆学研究, 2013 (6) .
- [12] 曹霞.国内 Data Curation 研究现状与热点分析 [J].图书情报工作, 2014 (18) .
- [13] 王芳, 慎金花.国外数据管护 (Data Curation) 研究与实践进展 [J].中国图书馆学报, 2014 (12) .
- [14] 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) .
- [15] 刘桂锋, 钱锦琳, 卢章平.国内外数据治理研究进展:内涵、要素、模型与框架 [J].图书情报工作, 2017 (21) .
- [16] Khatri V, Brown C V.Designing data governance [J].Communications of the ACM, 2010 (1) .
- [17] Martijn N, Hulstijn J, Bruijne M, et al. Determining the effects of data governance on the performance and compliance of enterprises in the logistics and retail sector [C]. International Federation for Information Processing, 2015.
- [18] 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 (18) .
- [19] 许晓东, 王锦华, 卞良, 孟倩.高等教育的数据治理研究 [J].高等工程教育研究, 2015 (5) .
- [20] 李青, 韩俊红.数据治理:提升教育数据质量的方法和途径 [J].中国远程教育, 2018 (8) .
- [21] 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 (5) .
- [22] 刘洁璇.高校图书馆知识管理中的数据治理 [J].情报科学, 2018 (1) .
- [23] 任亚忠.从数据管理走向数据治理——大数据环境下图书馆职能的转变 [J].四川图书馆学报, 2017 (4) .
- [作者简介] 付博 (1990-), 女, 内蒙古通辽人, 国家开放大学学习资源部 (数字图书馆) 科员, 助理馆员, 硕士, 研究方向: 数字图书馆建设与管理; 武兴华 (1990-), 男, 河北石家庄人, 国家开放大学信息化部科员, 研究实习员, 硕士, 研究方向: 开放教育教务管理。
- [责任编辑 时锦雯]