

大数据时代档案数据治理研究

金波 杨鹏

(上海大学图书情报档案系 上海 200444)

摘要 档案数据是国家经济和社会发展的战略性信息资源,在国家治理现代化、政府治理重塑、社会政策优化、公共服务提升等各项工作中发挥着基础信息支撑作用,开展档案数据治理研究具有重要的理论意义和实践价值。本文在分析档案数据治理研究动态的基础上,探析了档案数据治理内涵,提出从档案数据质量控制、整合集成、共享利用、安全保障与运行机制等方面构建大数据时代的档案数据治理体系,指导档案数据治理实践,实现档案数据善治。

关键词 大数据时代 档案数据 治理

DOI: 10.16065/j.cnki.issn1002-1620.2020.04.004

Research on Archival Data Governance in Big Data Era

JIN Bo, YANG Peng

(Department of Library Information and Archives, Shanghai University, Shanghai 200444)

Abstract: Archival data is a strategic information resource for national economic and social development. It plays a basic information supporting role in the modernization of national governance, reconstruction of government governance, optimization of social policies and promotion of public services and other work. It is of great theoretical and practical significance to carry out the research on archival data governance. Based on the analysis of the research trends of archival data governance, this paper analyzes the connotation of archival data governance, and puts forward the construction of archival data governance system in big data era from the aspects of archival data quality control, integration, sharing, security and operation mechanism, so as to guide the practice of archival data governance and realize the good governance of archival data.

Key words: Big data era; Archival data; Governance

0 引言

治理是现代社会、政治、经济、文化高度发展的产物。1995年,全球治理委员会在《我们的全球伙伴关系》研究报告中指出:“治理是各种公共的或私人的个人和机构管理共同事务的诸多方式的总和。它是使相互冲突的或不同的利益得以调和并且采取

联合行动的持续的过程。”^[1]2013年党的十八届三中全会指出:“全面深化改革的总目标是完善和发展中国特色社会主义制度,推进国家治理体系和治理能力现代化。”2017年党的十九大报告提出:“加强社会治理制度建设,完善党委领导、政府负责、社会协同、公众参与、法治保障的社会治理体制,提高社会治理社会化、法治化、智能化、专业化水平。”在全

面推进国家治理体系和治理能力现代化的背景下,我国各个领域、各个方面都在全面深化改革,强化治理体系和治理能力建设。随着信息技术与经济社会的融合渗透,全球数据呈现爆发增长、海量集聚的特点,“截至2016年,全球数据体量增长到16.1ZB(1ZB=1000EB),根据国际数据公司(International Data Corporation)发布的白皮书《数据时代2025》预测,到2025年,这一数字将扩展至163ZB,相当于2016年的10倍。”^[2]大数据时代,万物皆可量化,数据与物质、能量一起成为构成世界的三要素。数据的核心地位、核心价值和核心动力日益凸显,数据成为重要的战略资源和无形资产,强化数据治理已成为国际社会广泛共识。

档案事业作为国家治理体系的重要组成部分,国家管理方式的调整必然引起档案管理方式的变革。2008年第16届国际档案大会主题为“档案、治理与发展:映射未来社会”,推动档案界开始关注探索社会治理对档案工作的影响。2016年《全国档案事业发展“十三五”规划纲要》提出:“加快完善档案治理体系、提升档案治理能力,为夺取全面建成小康社会决胜阶段的伟大胜利作出积极贡献。”^[3]档案作为一种原生性信息资源,应在国家治理中发挥重要作用。大数据时代,档案数据大量形成,档案数据指数增长与管理落后、社会需求与保管分散、长期保存与载体脆弱、海量资源与价值实现等之间的矛盾日益凸显,“重创造轻管理、重数量轻质量、重保存轻利用”现象突出,数据污染、数据异构、数据安全、数据孤岛等问题普遍存在,造成档案数据“不可知、不可控、不可取、不可联、不可用”,给档案资源建设、服务创新、开放共享、安全合规、隐私保护等带来巨大挑战。究其根源,在于档案数据治理的缺失、滞后和失衡,难以适应档案事业发展需求与国家社会治理要求。因此,迫切需要加强档案数据治理。

1 档案数据治理研究动态

目前,档案治理研究主要集中在档案资源建设、档案法规建设、档案安全等方面,对档案数据治理研究相对较少,主要体现在以下三个方面:

一是档案数据治理动因研究。Woolfrey Lynn(2009)研究了调查数据档案在非洲地区治理中的作用,认为“调查数据档案是推动政策制定与社会研究的重要动力,非洲缺乏有效的调查数据档案,阻碍了国家政策的制定”。^[4]安小米(2015)认为,

“档案部门推行双套双轨制,制约了政府治理信息化能力的提升。”^[5]钱毅(2018)认为,“随着‘上游’部门的业务实践已经开始规模性地呈现出数据态的典型特征,数据(化)转型已经引发文档部门的‘生存’危机,催生数据管理问题。”^[6]于英香(2018)认为,“海量档案数据的管理、大数据时代智能数据管理、档案管理效率等推动着档案数据治理的产生。”^[7]

二是档案数据治理主体研究。Isabel Schellnack-Kelly等(2015)分析了档案工作者和文件管理人员在地方治理中的角色,“以南非为例指出治理、消除贫困等问题都依赖于档案工作者和文件管理人员对档案信息的保管和获取”。^[8]冯惠玲(2017)指出,“只有多方合力和联动,共同治理,才能有效推动文件管理数字转型的顺利进展。”^[9]刘越男(2018)对政务文件的归档主体地位进行了分析,提出“档案部门要借助既有的档案行政管理网络,通过制度、组织、机制、契约等治理技术,和相关机构结盟,协同开展政务文件归档。”^[10]

三是档案数据治理策略研究。澳大利亚国家档案馆(2016)以政府信息资产为核心制定了政府信息治理框架(Information Governance Framework),“将多元治理、绩效评估和风险管理等先进理念融入政府信息策略中,并将其贯彻落实;通过问卷调查形式定期核查政府机构信息治理策略的执行情况”。^[11]梁凯(2018)认为,“档案管理部门在数据治理体系下有计划地在风险应对、组织业务活动记录、保存计划、元数据收集、访问控制、格式管理、数字取证等方面建立规章制度、标准体系以及技术工具集合,融入政府数字化转型。”^[12]

从目前的研究来看,面临的问题主要有:一是理论研究深度不足,未能全面分析大数据时代档案数据治理面临的机遇和挑战,对档案数据治理的内涵、方式、运行机制、实现路径等缺乏深层次研究,档案数据治理理论有待探索建立;二是缺乏整体性思考,相关研究内容较为零散,不够系统完整,整体理论框架尚未建立,对大数据时代档案数据治理体系缺乏系统研究和整体认识。

由此可见,开展大数据时代档案数据治理研究有着巨大的理论潜力和现实需求。一是能够拓展档案学研究思路。将治理理论与档案数据关联起来,从治理理论审视档案数据管理,推动档案学理论与治理理论交叉渗透融合,拓展档案管理研究领域,促进档案数据资源建设,丰富档案学研究内涵,为档案

学研究提供新思路、新路径。二是开辟档案数据治理研究新领域。大数据时代数据资源的空间结构发生颠覆性改变,面对数据飞速膨胀、高度分散、价值多元、深度共享的特征,迫切需要开展档案数据治理研究,对档案数据治理主体、运行规律、内在关联、治理手段等进行多维度思考和深层次审视,形成档案数据治理研究范式。三是提升档案数据治理能力。通过创建大数据时代档案数据治理体系,着力探究档案数据治理运行机制及实现路径,有效提高档案数据管理效率与效能,增加档案数据公共服务供给,减少“数字鸿沟”和“数字不平等”,达到档案数据善治的目标。

2 档案数据治理内涵

2.1 档案数据治理环境

数据环境。大数据时代,档案数据的时空环境发生显著变化。从时间上看,“大数据时代的档案数据是传统档案数据和大数据时代产生的实时新数据之整合。”^[13]数量急剧增长,规模迅速膨胀,资源总量巨大,时效特征明显。全国而言,“调查显示,截至2018年,我国已有33.1%的国家综合档案馆数据存储量达到了1TB以上。”^[14]地方综合档案馆以青岛市为例,“截至2017年2月底,青岛市档案馆馆藏数据量已经超过25TB,而且还将以每年2TB的速度增加。”^[15]但与此同时,档案数据质量不合规、格式不统一、标准不规范、开发效率低等问题日益显现。从空间上看,大数据时代档案数据的空间范围不断延展,数据获取的范围和边界日益开放,数据的来源更加分散、结构更加复杂、类型更加多样、形式更加多元、管理难度更大、开发要求更高。大数据时代,档案数据时空环境的变化,迫使档案界必须高度重视对这些体量巨大、稍纵即逝、价值集聚的数据的治理。

政策环境。当前,国家、行业政策话语体系的大数据转向引领档案数据开展治理。国家层面,2015年国务院颁布《促进大数据发展行动纲要》,“对我国大数据发展进行国家层面的顶层设计和战略部署,赋予大数据作为建设数据强国、提升政府治理能力、推动经济转型升级的战略地位。”2018年颁布的国家标准《信息技术服务 治理 第5部分:数据治理规范》(GB34960.5—2018)提出了数据治理的总则和框架,规定了数据治理的顶层设计、数据治理域、数据治理过程和数据治理要求等。档案部门也紧跟国家重大战略步伐,2016年《全国档案事业发展

“十三五”规划纲要》提出:“制定档案数据开放计划,落实数据开放与维护的责任;优先推动与民生保障服务相关的档案数据开放;积极探索助力数字经济和社会治理创新的档案信息服务”^[16]。2019年国家档案局局长李明华提出,“聚焦有效服务国家治理体系和治理能力现代化、加快推进档案治理体系和治理能力现代化,进一步提高档案管理水平和服务能力”。^[17]总体而言,逐步完善的政策环境为治理的深入开展提供顶层设计、战略指导和方向引领。

社会环境。随着信息技术的广泛应用和经济社会的快速发展,社会公众精神文化需求更加多样,“数据对社会生活的广泛渗透与传播流动客观上赋予了公众包括数据权在内的个体权利”^[18],公众的个体数据意识、数据获取程度、数据利用能力不断提高。一方面,满足社会信息需求是档案服务的永恒主题,但当前档案数据开放程度不足,与公众密切相关的民生数据、舆情数据等相对匮乏,难以满足社会日趋精细化、个性化、多元化的档案数据利用需求;另一方面,档案查阅、档案编研、档案展览、参考咨询等传统档案服务方式难以满足大数据时代档案用户的网络化、便捷化、精准化、知识化和高效化的服务需求。大数据时代,档案服务需要由“供给导向”转向以用户为中心的“需求导向”。

技术环境。随着信息技术的迅猛发展,社会各项事业的信息化、网络化、智能化程度越来越高,信息科技为档案数据治理提供了强大的技术支撑和保障。大数据时代,档案数据具有海量增长、来源广泛、形式多样、价值集聚等特征,运用传统的技术方法难以进行有效管理,而大数据、云计算、人工智能、物联网、区块链等现代信息技术为档案数据管理与服务提供了可能。比如,依托大数据技术建立高效、便捷的档案数据处理模式;利用数据挖掘技术促进档案管理的精细化;借助人工智能、文本挖掘、语义分析等技术挖掘海量数据的潜在价值,将数据转化为知识;利用VR、AR、MR等大数据可视化技术将结构复杂、类型多样的档案数据以清晰明朗、直观生动的方式呈现出来,提高档案资源的利用率。

2.2 档案数据治理对象

档案数据治理以档案数据为对象,围绕档案数据收集、管理、存储、利用全生命周期展开,数量丰富、价值集聚的档案数据构成治理的基础性要素。档案数据有狭义和广义之分。狭义的档案数据是指国家机构、社会组织和个人在社会各项活动中形成的具有保存和利用价值,以数码形式记录于磁盘、光盘

等载体,依赖计算机系统存储、阅读和处理,并在通信网络上传输的各类电子文件及其元数据^[19]。广义的档案数据是指各级各类档案机构收集保存的具有档案性质的数据记录,包括各种数据形式的档案资源,如各类数字档案、多媒体档案;也包括档案管理与利用过程中产生的数据,如档案网站的浏览记录、平台日志、查阅服务数据、档案统计数据等。本文中的档案数据是指广义的档案数据。

档案数据具有以下几方面特点。一是来源广泛。大数据环境下,数据获取的范围和边界日益开放,“档案部门所收集的数据既包括传统数据资源,也包括大数据环境下产生的新型数据资源,如来自传感器、物联网的感知数据;来自微博、Web日志、点击流等的社交媒体交互数据”^[20];以及档案管理与利用过程中产生的各类数据。此外,还包括尚未纳入档案部门保管的数据,如政府信息公开数据、档案用户数据等。二是结构复杂。大数据环境下,原生数据、衍生数据大量存在,结构化、半结构化和非结构化数据混杂其中,如电子邮件、音频、视频、点击流、日志文档和生物计量学数据等结构复杂的档案数据。三是类型多样。档案数据既包括以往数字化档案、电子档案等历史累积数据,也包括当前形成的各类数据。根据数据产生平台的不同,可划分为政务档案数据、企业档案数据、社会档案数据等^[21];根据产生档案数据活动类型的不同,可划分为内容档案数据、管理档案数据和利用档案数据等。四是数量巨大。大数据时代,数据急剧增长,对于档案部门而言,随着档案数字化和接收电子档案的逐年增加,各档案馆(室)数据资源总量巨大。“截至2018年底,全国各级国家综合档案馆馆藏电子档案127.7万GB。其中,数码照片33.2万GB,数字录音、数字录像30.5万GB。馆藏档案数字化副本1556.4万GB。”^[22]同时,网络环境下,微信、微博等社交媒体产生的档案数据,各类网站产生的档案数据等规模也堪称海量。五是价值集聚。大数据时代,数据加速度增长,数据的处理速度更快、效率更高,数据实时归档,状态与价值随时空变化不断演变。档案数据作为价值密度高、关联意义大、决策依据性强的数据记录,是一种最真实、最可靠、最重要、最具权威性与凭证性、最具保存价值的原生数据资源,能够利用档案数据提供实时、准时、即时的数据分析与决策服务。

2.3 档案数据治理定义

档案数据不仅是一种重要的数据记录,也是重要的社会记忆,是国家经济和社会发展重要的战略

性信息资源,在国家治理现代化、政府治理重塑、社会政策优化、公共服务提升等各项工作中发挥着基础信息支撑作用。档案数据治理是指档案部门、社会组织和公民等多元主体协同合作,依据一定的法规标准,充分利用大数据等现代信息技术,对档案数据生成、收集、管理、存储、利用整个生命周期进行科学规范的全程管理,挖掘档案数据价值,满足社会利用需求,实现档案数据善治的活动和过程。档案数据治理需要坚持协同共治、互联互通、全程治理、风险可控等原则,推动档案数据治理主体协同化、治理决策科学化、治理流程标准化、治理手段智能化、治理过程透明化、治理技术迭代化、治理结果精准化^[23],促进多元主体互信互任互动、治理方式合法合规合约,实现数据全生命期、全流程、全要素的综合集成管理。档案数据治理围绕档案数据展开,形成以档案数据为资源、为基础的网络化、扁平化、共享化、平台化的数据型生态环境,推动档案价值体现、资源禀赋创新、社会秩序重构、国家治理变革。档案数据治理需要利用各种现代信息技术、严密有效的管理制度,形成多元共治、精准治理的治理架构,构建以档案部门为主导,多元主体协同治理,融合制度、管理、技术于一体的动态治理体系。档案数据治理的目标是实现档案数据善治,即确保档案数据真实完整、质量可靠、安全可控、运行合规和价值实现。

档案数据治理是数据治理与档案治理的交叉契合,三者之间关联紧密、互为支撑。根据国家标准《信息技术服务治理 第5部分:数据治理规范》,数据治理是指数据资源及其应用过程中相关管控活动、绩效和风险管理的集合。数据治理源于组织的外部监督、内部数据管理及应用的需求。数据治理的目标是组织通过评估、指导和监督的方法,按照统筹和规划、构建和运行、监控和评价以及改进和优化的过程,保障数据及其应用过程中的运营合规、风险可控和价值实现。档案数据作为国家和社会重要的数据资源,与数据治理存在天然契合。档案数据治理是数据治理的重要子集,二者在数据质量把控、资源整合集成、共享平台建设、风险安全防范、价值深度挖掘等方面存在相似性,档案数据治理可以借鉴数据治理已有的经验、技术、工具以及方法,创建档案数据治理理论体系。档案治理是指“档案部门、社会组织和公民个人等多个主体协同合作,基于一定的行动规则,共同对档案事务进行科学、规范管理,实现档案领域善治的活动和过程。”^[24]“档案治理以档案管理

为基础,是治理理论在档案领域的应用,强调档案部门、社会组织、市场和公民协同共治,是促进档案事业发展的重要方式。”^[25]档案数据治理是档案治理的重要组成部分,构建大数据时代档案数据治理体系,有利于推动档案管理走向档案治理,对完善档案治理体系、提升档案治理能力和水平、促进档案治理现代化具有重要意义。

2.4 档案数据治理特征

数据驱动。大数据推动生产要素的集约化整合、网络化共享、协作化开发和高效化利用,数据已经成为社会各个领域业务创新增值、提升核心价值、推动产业变革的重要引擎,是推动社会治理有序展开的重要驱动力。数据的广泛渗透与流动传播,导致众多行业的业务形态向数据态转变,体量巨大、来源广泛、结构复杂的档案数据开始成为档案工作的对象。档案数据作为高价值、信任度强的记录凭证,利用档案数据能够增强政府决策的科学性、合理性、准确性,提供优质的公共产品和公共服务,建设服务型政府;能够提高社会参与度,推动社会治理多元协作、高效智能、结果精准,提升社会治理公信力;能够推动打破“信息孤岛”和业务藩篱,构建内外协同、上下联动、纵横互通的整体性治理大格局。档案工作对象的变化推动档案管理方式的变革,数据关联意义强、数据共享程度高、数据管理空间广的液化管理方式成为档案管理的新方式。在此情境下,数据驱动乃至模型驱动型业务系统普遍出现,将不同来源、不同系统、不同业务环境且相互关联的数据集成到一起,形成各类专门的数据体、数据湖、数据球、数据大脑等。随着“前端”业务部门大量数据体的产生,驱使作为“后端”的档案部门必须拓展管理范围,重新审视如何对数据体进行“档案化”管理。

协同联动。在传统的档案数据管理模式下,档案数据呈碎片化、分散化,不同地区、不同层级的档案部门各自为政、协同不足,“数据孤岛”和“数据打架”现象并存,档案数据共享成本高、效率低。通过档案数据治理,发挥数据潜在的系统性、流动性、共享性特征,强化档案部门与业务部门、文书部门、信息部门、大数据部门等之间的协同联动、多元共治;以档案数据集中和共享为途径,推动技术融合、业务融合、系统融合,打通数据壁垒,弥合不同层级和部门之间的鸿沟,重构碎片化的治理格局,形成覆盖全国、统筹利用、统一接入的档案数据共享大平台;加强档案数据的整合与融合,构建全国档案数据共享

体系,实现跨层级、跨地域、跨系统、跨部门、跨业务的档案数据协同管理和服务^[26],形成档案数据治理的强大合力。

精准高效。大数据时代,数据的时效性特点更加明显,处理方式更多、速度更快、效率更高,档案部门需要借助数据分析、数据关联、数据挖掘技术,深入了解公众需求,对社会需求进行多维度、多层次细分,实时、全面、准确的预测和发现社会公众的需求热点;建构档案数据库,实现基于海量档案数据的识别、分析与匹配,发挥高效治理手段的赋能,提供及时、快速、优质的档案数据服务;优化档案业务流程,增强决策前瞻性,提升运作效率,盘活数据价值,为公众提供更有针对性,更加个性化、精准化、一站化、知识化、智能化的档案数据服务,提升治理绩效。如“2014年,上海建成世界上最大的电子健康档案信息库和PB级医学影像档案库,基于两库启动‘医联工程’,为每一位病人自动生成电子健康档案,依据病人的健康数据,医生能够做出更加精准的诊断;病人根据数据库的提示可以做出更恰当的就医选择;政府则可以提高对疾病和疫情的处置效率和能力。”^[27]

安全为重。安全是档案工作的第一要务,数据是档案数据治理的“生命线”。2016年《“十三五”国家信息化规划》提出,“注重数据安全保护。实施大数据安全保障工程,加强数据资源在采集、传输、存储、使用和开放等环节的安全保护。”^[28]2017年李明华在“全国档案安全工作会议”上指出,要以维护档案安全为中心,“确保档案数据、信息系统及网络始终可用可控”^[29]。当前,档案数据存在信息伦理、知识产权保护、隐私侵犯、系统平台漏洞、技术更新换代、黑客病毒侵袭、载体寿命不确定等各种安全风险,需要充分利用各种技术、管理和制度,加强档案数据安全治理,促进档案数据安全风险预测化、管控精准化、防御纵深化、应对即时化,确保档案数据安全存储和利用。

3 档案数据治理体系

大数据时代,档案数据治理应以档案数据质量控制为基础,构建档案数据质量标准新规范;以数据整合集成为核心,探索档案数据资源建设新范式;以共享利用为目标,创新档案数据社会共享与公共服务模式;以档案安全为基石,构筑档案数据安全保障新途径;以公共参与为导向,推进多元、协同、融

合为一体的档案数据治理新机制,建立科学合理的档案数据治理体系,指导档案数据治理实践,实现档案数据善治。

3.1 档案数据质量控制

数据质量是指“数据的一组固有属性满足数据消费者要求的程度,可以通过数据的完整性、及时性、有效性等一系列特征属性的集合来描述”。^[30]数据质量是档案数据治理的关键,是档案数据规范管理与有效开发的基础。档案数据质量包括两方面:一是档案数据自身的质量,主要指档案数据的真实完整、准确可靠、关联可用程度等;二是档案数据的过程质量,包括数据存储质量、传输质量、运行质量和利用质量等。

评估档案数据质量主要有以下几个维度:完整性,指档案数据保持未分割、未破坏、统一整体的状态,数据实体、数据属性等是否缺失;准确性,指档案数据是否与其记录的事实、实体一致,即档案数据的真实权威程度;规范性,指档案数据的数据格式、数据类型、值域对相关标准规则的符合程度;一致性,主要指随时间变化,档案数据是否存在变异或变更;关联性,档案数据不是孤立存在的,数据之间往往存在着各种关联;可控性,指档案数据在运行、管理、利用过程中可以控制的有效度。导致档案数据产生质量问题的因素主要包括:信息因素,主要指元数据描述错误、业务规则和校验规则错误、数据集成不当等;技术因素,主要指档案数据处理中的各种技术异常造成的数据质量问题;流程因素,主要指由于在档案数据管理过程中,系统作业流程和操作流程设置不当造成的数据质量问题;管理因素,是指由于管理机制、管理人员等方面造成的数据质量问题。当前,档案数据来源广泛、形态多样、结构复杂、格式不一、系统异构、标准差异等现象普遍,档案数据质量不容乐观,需要立足档案数据管理现状,加强档案数据质量控制。

档案数据质量控制是指针对档案数据全生命周期流程中的各种数据质量问题,进行识别、度量、监控、检测、分析、回溯、预警、解决、反馈、评估等一系列管理活动,从而改善和提高档案数据质量的过程。档案部门应根据国际标准ISO9000质量管理体系,加强档案数据标准制定,依据标准对档案数据进行分类管理;明确各主体在档案数据质量管理中的角色和职责,建立档案数据质量管理体系及实施机制,研发档案数据质量控制技术;全面分析档案数据的生成、流转、归档、保管、存储、利用全生命周期

流程,准确把控档案数据质量影响要素,创新档案数据质量管理方法,构建档案数据质量评估框架;从认知、制度、技术等方面完善涵盖质量标准、控制方法、清洗规则和评价体系等在内的档案数据质量控制规范体系,保障大数据时代档案数据资源建设质量,满足档案数据质量应用需求,为档案数据资源有效利用与共建共享奠定基础。

3.2 档案数据整合集成

当前,档案数据资源管理无序与有序、分散与集成、孤立与互通、异构与统一之间的矛盾突出,需要通过建立有效的档案数据资源整合策略、整合模式和整合方法,力争将离散、孤立、碎化、多元、异构的档案数据资源通过逻辑方式或物理方式组织成一个有机整体,填平不同地区、不同层级、不同部门之间的“数字鸿沟”。

首先,要加强档案部门间的档案数据整合。档案行业“条块分割”的体制壁垒、各自为政的管理方式、行业部门间的数据差异等给档案数据整合集成带来困难。档案部门要加强数据治理规划,按照“统一平台、互联互通,存量共享、增量共建,物理分散、逻辑集中”的原则,利用大数据技术,打破部门隔阂和“条块分割”的碎片化状态,针对不同来源的档案数据,统一数据格式,制定元数据标准与政策,构建数据标识,实施元数据管理;汇集和统筹数据资源,加强馆内、馆际数据资源融合和聚集,弥合部门之间、业务之间的“数据缝隙”,形成“数据集”“数据链”“数据湖”“数据仓”“数据群”“数据云”等统一的档案数据资源平台,彻底打破档案部门“信息孤岛”的状态。如杭州市档案局在“最多跑一次”工作的引领下,“针对电子业务数据归档,开发了一个能够集成多系统多数据源的数据集成平台——杭州市电子业务数据归档系统。通过整合各类业务数据,为各类业务系统提供业务数据的长期保存、管理和利用,缓解了各类机关单位信息系统面临的数据膨胀压力,扩展了档案部门保存的数据种类和数据量,提升了档案部门的数据管理和服务能力。”^[31]

其次,加强档案部门与信息部门、政府数据管理部门间的档案数据整合。当前,大数据局、大数据管理中心等各类大数据管理机构涌现,如上海市大数据中心、天津市大数据管理中心、贵州省大数据管理局、杭州市数据资源局等,均承担政务数据的归集和应用融合工作。面对复杂的业务、系统和数据环境,档案部门尚无法全面收集、整合档案数据,给作为档

案资源保管存储机构的档案部门带来了巨大冲击。档案部门需要主动出击,与信息部门、数据管理部门加强协作,共同整合档案数据,构建统一的档案数据资源管理平台。如湘潭市档案局从2011年开始积极对全市数据资源管理进行探索,把数据资源体系建设提升到战略层面。“2012年6月,在湘潭市政府协调下,以湘潭市档案局为主导,整合各类数据管理平台、部门,在全国率先成立档案数据中心。”^[32]

最后,加强档案部门与社会间的档案数据整合,形成档案数据资源群。档案部门需要借助数据集成、交换、仓储、挖掘等大数据技术,整合公共领域企业、社会组织、民众所拥有的民意社情数据、安全保障数据、公众健康数据、防灾减灾数据等,整合来源于不同渠道的公共网络数据、公众民生数据、物理空间数据等,保障社会产生的档案数据畅通融入档案数据治理的数据仓库中;建立具有丰富馆藏的集成性档案生态数据库,引导激励公众参与档案数据资源提供,推动社会档案数据资源集成管理,形成档案数据集成平台,更好地满足社会利用需求。

3.3 档案数据共享利用

大数据时代档案用户的利用需求更加多样化,需要立足社会信息资源共建共治共享理念,加强档案数据公共服务转型与创新,促进档案数据资源的深度开发及有效利用,构建精细、精简、精准、智慧的档案数据公共服务体系,推动档案数据服务的智能化、社会化、个性化,把“资源库”变成“知识库”“思想库”“智库”。

首先,档案部门要制定档案数据开放政策和制度,优先推动民生档案数据开放,提供开放档案数据资源的社会化服务;其次,建立国家档案数据库,整合各类数据资源,构建统一的档案数据共享平台,提供档案数据的一站式服务;再次,要利用数据关联分析、数据挖掘、人工智能等技术,对用户身份、借阅内容、搜索记录等数据的相关性、因果性进行关联分析,实现无序数据向关联化转化、海量数据向智能化转化、静态数据向动态化转化、隐性数据向显性化转化;最后,要深度挖掘档案数据价值,创新档案服务模式,深化档案服务内涵,预测用户隐性诉求,提供个性化、关联化、精准化、跟踪化、人性化和知识化的档案服务,提升档案服务能级。对此,中国联通电子档案数据挖掘与智能利用的经验值得借鉴。“大数据时代,企业档案数据呈现海量趋势。传统的档案服务方式,难以应对日益增长的数据需求。对此,中国联通电子档案馆整合档案信息资源库

和档案知识库,构建档案数据挖掘服务平台、多维展示平台和知识图谱系统,为企业智能化管理和决策提供支撑。”^[33]

3.4 档案数据安全保障

大数据环境下,档案数据在收集、存储、管理、利用过程中存在各种安全隐患,面临的安全风险主要有4项。一是档案数据安全法规体系不健全。针对档案数据安全的相关法律尚不完整、不配套、不健全,有关档案数据安全问题的行业、地方标准规范有待完善。如档案数据在转移、共享过程中涉及的知识产权问题尚无明确的安全责任界定。由于传统档案具有原始性或“孤本”的概念,因此在档案转移过程中一般不存在所有权、处置权和使用权相分离的情况。而档案数据则不同,由于档案数据内容信息和特定载体的可分离性,以及内容信息的易复制性和易修改性,因而不存在传统意义上的“原件”概念。对于档案数据,某人的所有、处置和使用可以不影响其他人的所有、处置和使用,三种权限可以相分离,由此使档案数据在归档移交时面临复杂的权属关系和处置情况,而当前尚无明确的法规条文予以规定。二是档案数据面临被黑客攻击和利用的风险,黑客借助于网络攻击、系统漏洞、病毒传播等方式,造成档案数据被篡改、窃取、泄露和贩卖。三是针对档案数据在存储、传播、利用等运行过程中的各种安全风险,档案部门缺乏必要的安全管理知识、安全控制技能,普通民众对档案数据的认知存在不足,缺乏基本的数据安全素养,档案数据安全意识较薄弱,需要进一步加强档案数据安全宣传,健全档案数据安全分级、保密、备份、应急、监控等安全管理制度。四是档案数据的载体安全,即档案数据媒介的寿命问题、载体的兼容问题和可读取问题等。

首先,要加强档案数据安全领域的立法和标准建设,加强宏观治理,建立完备的档案数据安全法规体系,清晰地界定档案数据权属和各方安全责任,完善档案数据隐私保护的相关立法,为档案数据安全体系建设提供法律支撑;其次,要完善档案数据安全分级、保密、备份、应急、监控等管理制度,从制度层面保障档案数据安全;最后,要构筑档案数据安全技术壁垒,利用入侵检测技术、加密技术、区块链技术等保障档案数据安全,实现安全技术自主可控。一方面,可以防止数据窃取、泄密和破坏,实现档案数据安全保密;另一方面,能够防止数据丢失和损毁,确保档案数据真实、完整和可读。如“苏州大学苏航档案数据保全中心通过采用第三方国家

时间戳授时中心签发的电子证书,在不获取档案数据内容的情况下证明档案数据的存在性和内容完整性,确保档案数据长久保存的真实性、安全性、可用性。”^[34]

3.5 档案数据治理运行机制

国家治理要求政府放权,推动档案管理转型发展,促进档案管理走向档案治理。运行机制关系到档案数据治理的成效,需要系统建立大数据时代档案数据治理的运行机制,推进档案数据治理社会化、专业化、现代化、智能化。档案数据治理涵盖“多元”“互动”“融合”等诸多新要素,需要以大数据时代档案工作数字转型为契机,采用档案部门“元治”、多主体协同治理、多手段融合治理等运行机制,创新档案数据治理方式,优化档案数据治理生态,提高档案数据治理能力。

档案部门“元治”。档案部门在档案数据治理中担负着“元治理”的重要角色,是档案数据治理的主要承担者,在档案数据治理中要发挥主导作用。一方面,档案部门是档案数据的聚集地,应主导档案数据公共服务体系建设;另一方面,档案数据治理主体多元、关系复杂,当治理主体间出现矛盾纷争、难以协调时,就需要发挥档案部门的“元治理”作用。因此,档案部门应树立“元治”理念,强化档案数据治理职能。首先,档案部门要树立数据全域观,完善档案数据治理法律法规制度,为档案数据协同共治营造良好生态,为社会主体积极参与档案数据治理创造条件;其次,档案部门要重新思考档案数据收、存、管、用方式,探究档案数据生成环境、形成规律、构成要素、运行过程,强化数据本体研究;最后,档案部门要加强大数据理论与技术学习,加快专业人才引进和基础设施建设,综合运用政策、法规、标准、技术、工具等,通过更为专业、精深、智能的治理手段,完善档案数据服务利用体系。

多主体协同治理。依据治理理论,档案数据治理应秉持多元共治理念,形成多元化、多层次、多维度的立体协同治理模式,强化档案部门与政府、社会、公众等多主体协同合作,形成纵横交错、互联互通、互助共赢的治理协作机制,促进档案部门与社会之间的深度融合。推进档案数据治理,档案部门应明确其处于主导而不是垄断、服务而不是掌舵的角色,摒弃将社会组织与民众视为单纯的档案消费者的传统管理观念,平等对待社会主体,积极培育扶持社会力量参与档案数据治理的决策与管理,通过分权和赋权明确其角色和定位,构建分工协作、共同参与的

档案数据治理新格局。社会组织、商业机构、公众等其他主体需要发挥自身专业、技术、资源优势,积极主动参与档案数据治理。如澳大利亚在政府档案数据治理过程中,形成了政府部门主导、社会组织、公众等多元主体协同治理的良性格局。澳大利亚国家档案馆负责收集、管理、公开联邦档案数据,保障档案数据安全;社会组织通过协助档案部门制定数据政策、提供技术支撑、参与国际化合作等方式主动参与治理;社会公众则通过数据利用、数据上传、数据监督等形式积极参与治理。

多手段融合治理。档案数据治理需要综合运用行政、法治、市场、社会等多种治理手段,通过理念、标准、制度、技术、工具、流程、人才等治理要素多管齐下,着力构建整体联动、相互协调的治理环境,促进主体联盟、过程联通、要素连接,灵活有效应对大数据情境下复杂多变的社会环境,创建良好的“建数”“管数”“用数”生态,实现档案数据科学规范管理;要创新档案数据治理方式,利用互联网扁平化、交互式、快捷性等优势,借助大数据等现代信息技术,通过市场运作、众筹众包、众扶众建、档案志愿服务等方式加强档案数据治理,弥补档案部门管理人员、技术力量、专业知识面等方面的不足和缺陷;引入市场化方式和社会力量开展档案数据治理,推动档案数据线下线上的融合治理、静态动态的双重治理、现实虚拟的综合治理。“2012年,瑞典美狄亚·马尔默大学发起‘Living Archives’项目,通过众包整活档案资源。公众可通过社交媒体查看、修改、注释档案资料,也可将收集的数据加入数据库中。”^[35]

舍恩伯格在《大数据时代》一书中认为大数据是思维革命和社会革命的标志。大数据时代,档案数据急剧增长,档案数据保存状态与档案数据治理水平不相匹配,只有不断地提高档案数据治理能力和治理水平,通过多层次互联、多维度互通、多学科互融构建全新的档案数据治理体系,才能充分挖掘档案数据价值,促进档案数据治理现代化,实现档案数据善治。

本文为国家社科基金重点项目“大数据时代档案数据治理研究”(19ATQ007)研究成果之一。

参 考 文 献

- [1] 俞可平.全球治理引论[J].马克思主义与现实,2002(1):20-32.

- [2] 王兆君,王钺,曹朝辉.主数据驱动的数据治理——原理、技术与实践[M].北京:清华大学出版社,2019:3.
- [3][16] 国家档案局.《全国档案事业发展“十三五”规划纲要》(档发[2016]4号)[Z].
- [4] Woolfrey Lynn.Knowledge Utilization for Governance in Africa: Evidence-based decision-making and the role of survey data archives in the region[J].INFORMATION DEVELOPMENT,2009(1):22-32.
- [5] 安小米,白文琳.云治理时代的政务数据管理转型——当前我国档案事业发展的的问题与建议[J].人民论坛·学术前沿,2015(16):72-84.
- [6] 钱毅.从“数字化”到“数据化”——新技术环境下文件管理若干问题再认识[J].档案学通讯,2018(5):42-45.
- [7][13] 于英香.从数据与信息关系演化看档案数据概念的发展[J].情报杂志,2018(11):150-155.
- [8] Isabel Schellnack-Kelly, Thomas B. Van der Walt.Role of Public Archivists and Records Managers in Governance and Local Development under a Democratic Dispensation [J].ESARBICA Journal,2015(34):1-20.
- [9] 冯惠玲,刘越男,马林青.文件管理的数字转型:关键要素识别与推进策略分析[J].档案学通讯,2017(3):4-11.
- [10] 刘越男.大数据政策背景下政务文件归档面临的挑战[J].档案学研究,2018(2):107-114.
- [11] National Archives of Australia.Digital Continuity 2020 Policy (October 2015) [EB/OL].[2020-2-6].http://www.naa.gov.au/Images/Digital-Continuity-2020-Policy_tcm16-93933.pdf
- [12] 梁凯.多维度视角下大数据与档案关系的思考——以杭州市大数据管理为背景[J].中国档案,2018(7):66-67.
- [14] 谢永宪,王巧玲,房小可,薛怡.我国国家综合档案馆数字档案信息长期保存现状调查[J].档案学通讯,2019(4):58-62.
- [15] 王金玲.档案数据的智慧管理与应用研究[J].中国档案,2018(4):61-63.
- [17] 李明华.在全国档案工作暨表彰先进会议上的讲话[EB/OL].[2020-02-20].http://www.saac.gov.cn/daj/yaow/202001/afbf92881b5c4f36a311316d1e3690da.shtml
- [18] 夏义堃.试论政府数据治理的内涵、生成背景与主要问题[J].图书情报工作,2018(9):21-27.
- [19] 湘潭市人民政府办公室.湘潭市人民政府办公室关于印发湘潭市档案数据管理试行办法的通知[EB/OL].[2020-02-20].http://www.xiangtan.gov.cn/1085/1082/1083/content_82798.html
- [20] 吴金红,张飞,鞠秀芳.大数据:企业竞争情报的机遇、挑战及对策研究[J].情报杂志,2013(1):5-9.
- [21] 郑金月.关于档案与大数据关系问题的思辨[J].档案学研究,2016(6):37-40.
- [22] 政策法规研究司.2018年度全国档案行政管理部门和档案馆基本情况摘要(二)[EB/OL].[2020-02-20].http://www.saac.gov.cn/daj/zhdt/201909/5482cc6929cd4b94b16a135517316d4f.shtml
- [23][27] 石火学,潘晨.大数据驱动政府治理变革[J].电子政务,2018(12):112-120.
- [24] 晏秦.论档案治理的内涵、特征和功能[J].档案管理,2017(4):4-7.
- [25] 金波,晏秦.从档案管理走向档案治理[J].档案学研究,2019(1):46-55.
- [26] 新华网.习近平:实施国家大数据战略加快建设数字中国[EB/OL].[2020-02-24].http://www.xinhuanet.com/politics/2017-12/09/c_1122084706.htm
- [28] 国务院印发《“十三五”国家信息化规划》[J].电子政务,2017(1):40.
- [29] 李明华.在全国档案安全工作会议上的讲话[J].中国档案,2017(7):14-21.
- [30] 熊兴江,马敬东,徐承中.电子健康档案数据质量评估与治理的综述研究[J].中国卫生信息管理杂志,2018(6):637-642.
- [31] 梁凯.“最多跑一次”事项电子业务数据归档系统建设实践——以杭州市档案局为例[J].浙江档案,2017(8):31-32.
- [32] 胡建略.从“管档案”到“管数据”[N].中国档案报,2013-11-08(002).
- [33] 杨茜雅.中国联通电子档案数据挖掘与智能利用的研究[J].档案学研究,2018(6):105-109.
- [34] 苏大苏航档案数据保全有限公司.档案数据保全中心[EB/OL].[2020-2-5].http://www.shdafw.com/shsj/security.asp
- [35] 刘磊,安小米,钟文睿.Living Archives项目及其对数字档案资源整合与服务的启示[J].浙江档案,2014(8):5-8.