

# 大数据背景下面向用户的档案数字资源建设研究

刘 蓉

航空工业计算所 陕西 西安 710000

**摘 要:** 面对大数据时代的到来,传统的档案建设方式已经不能够满足档案用户个性化、精准化的需求。本文通过对的文献研究总结出当前我国档案建设工作存在的不足与问题,并提出构建档案数字资源精准建设的模型。该模型借助大数据技术的支持,通过用户需求采集分析和档案精准建设两个系统实现以用户为中心的主动、精准、高效的档案建设工作,同时实现了用户数据采集、构建用户画像、提升档案建设水平等功能。为了保障模型顺利实施,本文从制度、档案馆自身建设、技术等多个方面提出措施建议,促进我国档案工作的与时俱进。

**关键词:** 档案数字资源建设;精准定位;大数据;档案用户需求

## 引言

随着当代互联网技术的发展和“5G时代”的到来,大数据在档案数字资源建设领域的作用变得越来越重要,展示出了巨大的发展潜力和空间。涂子沛在其著作《大数据:正在到来的数据革命》中提到:“大数据”之“大”,更多的意义在于:人类可以“分析和使用”的数据在大量增加,通过这些数据的交换、整合和分析,人类可以发现新的知识,创造新的价值,带来“大知识”、“大科技”、“大利润”和“大发展”。实际上,许多商业和情报服务业纷纷利用大数据技术转变传统的工作方式,以用户为中心提供主动的个性化服务,而档案工作部门对档案数字资源的建设方向仍然停留在传统的坐等用户上门利用,造成许多丰富的档案数字资源无法被用户充分了解使用,形成资源浪费。面对越来越发达的大数据技术,档案工作部门亟需转变档案数字资源建设的方式,立足于用户需求,通过构建用户画像实现精准化的数字资源建设工作,主动提供个性化的档案信息服务。<sup>[1]</sup>

## 1 研究现状与问题分析

### 1.1 研究现状

#### 1.1.1 大数据背景下的档案数字资源建设

大数据时代的到来给档案数字资源建设工作带来了挑战和机遇,档案界的许多学者也已经提出了大数据时代档案数字资源建设工作改进的方向;王芳学者基于大数据的背景下,提出了当前档案数字资源建设存在的问题,从制度和技术两种路径提出了档案数字资源建设工作的改进方向,对档案管理工作进行创新。崔艳红学者从小数据的角度出发对档案信息集成服务框架设计进行了研究,其框架围绕小数据管理系统和档案信息集成系统展开,以用户的小数据形成的用户需求为档案信息集成系统提供资源,从而实现对档案用户的个性化服务。

檀竹茂学者强调了协同合作对于开发档案数字资源的重要性,档案部门与社会力量实现协同合作,共同助力开发工作才能解决当前存在的问题。

#### 1.1.2 档案用户需求

伴随着大数据时代的到来和互联网的技术的发展对用户利用行为、利用习惯以及用户画像的研究提出了更高的要求,其中吕元智学者认为,用户行为是其信息需求的外在体现,从用户档案利用行为的视角来分析用户需求,形成一套系统的档案知识集成服务实现框架,实现个性化服务。连志英学者通过对档案利用需求进行调研,以中国近现代史学者为主体,对调研学者的用户背景信息和有关档案的查询利用情况记录两大部分进行调查研究,提出了档案信息服务工作的前进方向和不足。杨姝君学者则是对档案用户的利用需求进行了分类解释,总结出了档案用户利用需求规律,对档案部门如何更好满足用户需求提出了建议。关于对档案用户画像的研究,吕元智学者提出了基于用户场景开展档案服务工作的模式框架,以“用户小数据系统”“网络社交环境”作为需求分析系统的立足点,对用户的个性化需求习惯进行研究。

## 1.2 问题分析

### 1.2.1 档案数字资源建设的技术支撑薄弱

国家档案局原局长杨冬权认为,“数据=档案=财富”,数据越多,就越智慧,要树立合作开展档案数据化和数据挖掘的观念,积极推动档案的数据化发展。大数据时代的到来带来了数据的爆炸性增长和先进的数据处理技术,要实现对日益庞杂的档案数字资源的迅速、高效处理,就必须运用知识图谱、云计算、机器学习等大数据时代下产生的新技术来作为档案数字资源建设工作的支撑。

但是我国目前的数字档案建设工作仍然停留在原始的数据处理技术架构之下,数字档案馆建设中对大数据

技术的使用较少,面对日益庞杂的档案数据,数据的采集、存储、处理、利用等工作不能适应新的数据时代的要求,同时在平台构建、体系建设等方面,技术的缺乏导致缺乏统一的数据处理平台,或者平台的处理能力比较有限,阻碍档案数字资源建设开发的进程。

### 1.2.2 用户需求层面

档案数字资源的建设本质上是一项面向用户,为档案用户提供服务的工作,应当以用户需求为导向来选定数字资源建设工作的方向。

在档案编研主题的选择确定过程中,国家重大政策和社会重大热点焦点是公共档案馆开展工作的主要依据,用户需求则摆在了次要的位置上。在开展编研工作之前,也不会对于用户的需求进行专门调查,这与大数据时代下以用户需求为中心的档案数字资源建设工作有所违背,作为一项面向用户,为用户提供利用服务的工作,对用户需求和用户背景信息的把握是不可或缺的,而当前我国档案数字资源建设工作的开展之前对用户需求没有进行及时的收集总结,导致档案资源建设的成果与用户需求存在偏差,无法满足档案用户的需求,降低了档案数字资源建设的效率与成效。<sup>[2]</sup>

## 2 模型构建

笔者构建了大数据背景下对以用户为中心开展档案数字资源建设工作的模型,如图1所示。该模型主要包括两部分:用户需求信息采集分析系统和档案数字资源精准建设系统。两个系统之间相互影响作用,由大数据技术全面介入支持,实现对用户信息需求全面及时准确的收集,经过分析处理,形成用户标签体系,提取用户特征,构建用户画像从而进行档案数字资源的精准建设工作。

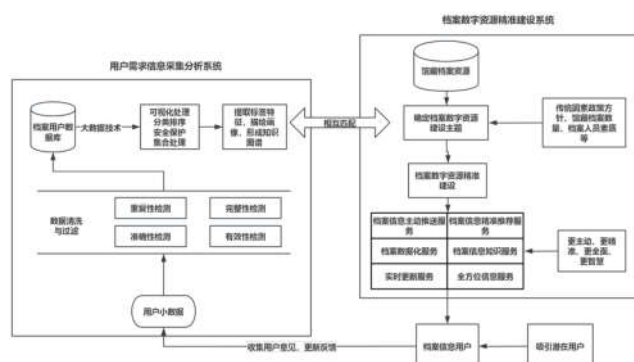


图1 面向用户的档案数字资源建设模型

### 2.1 用户需求采集分析系统

在大数据时代的新背景下,用户信息变得更加复杂多变,越来越多的用户通过移动终端进行档案信息的获取,因此对于用户所处场景的研究也也纳入到用户信息采集的工作范围内,是档案个性化服务工作的标志之

一。将采集来的一系列用户数据经过重复性、准确性、完整性、有效性的检测过滤之后就可以存放入用户需求的数据库之中,利用大数据技术对信息进行分析处理,提取用户特征,构建用户标签体系,形成档案用户画像来为档案数字资源精准建设系统提供智力支持。

### 2.2 档案数字资源精准建设系统

在现有馆藏档案数字资源的基础上,以用户需求分析系统提供的用户画像为中心,结合其他客观因素条件,实现对档案数字资源建设的精准定位,借助大数据技术的帮助,摆脱物理、时间、手段等方面的限制,更加主动、精准、个性、智慧的服务方式满足现有用户需求,同时通过良好的用户体验吸引更多的档案用户,反馈给用户信息采集环节更新用户数据。

### 2.3 模型实现功能

#### 2.3.1 实现高效全面的用户数据采集

在大数据技术的支持下,除了传统的用户背景信息外,增加对用户行为数据和用户场景等信息的调查研究,形成档案用户小数据概念。与大数据主要聚焦于宏观层面不同,小数据注重对每一位用户需求的记录研究,更好地适应了档案个性化服务这一目标的需求,实现高效率的用户信息采集工作。

#### 2.3.2 借助大数据技术构建精准用户画像

大数据时代除了带来更庞杂的数据集合,档案用户也越来越注重其自身问题的解决和个性化的档案信息需求,传统的档案用户需求简单分类已经不能适应工作的需要。另一方面伴随着数字档案馆、智慧档案馆等新形式档案馆的兴起,借助信息技术进行用户画像成为档案服务工作进步的必然要求。

采集来的用户数据通过筛选过滤之后存放在档案用户数据库中,档案工作人员可以利用大数据技术对用户信息进行标签化处理,挖掘不同用户之间的对应关系,构建个体用户和群体用户画像,借助数据可视化技术更全面地进行展现,来为用户提供更契合需求的信息服务。

#### 2.3.3 进行以用户需求为中心的档案建设

目前我国实体档案馆对于档案编研工作因素的考量主要集中于现有馆藏档案资源和国家政策支持等方面,用户信息需求因素没有得到足够的重视。而伴随大数据时代的到来,用户会产生更加多样化个性化的信息需求。借助用户需求分析工作产生的用户画像和知识图谱进行档案信息建设工作,能够最有效地契合用户的信息需求,从群体用户和个体用户两方面全面满足不同用户的信息需求。

## 3 保障措施

### 3.1 制度支撑

#### 3.1.1 用户反馈机制

在可视的用户交互界面增设用户的反馈页面能更加迅速便捷的获取用户的取向意见,档案部门通过对意见和反馈进行整理后,对档案数字资源的建设方式和内容更及时准确地做出调整来满足用户需求。

#### 3.1.2 监督机制

在读取用户信息和需求进行数据分析的过程中要有监督机制进行制约,应当由上级部门委派具有专业档案和法规知识的专门工作人员进行全程监管,严格按照法律政策开展工作,严格规定不同部门的工作人员所需要的访问和修改权限,保证信息安全和用户隐私不被泄漏。

#### 3.1.3 宣传机制

档案部门要借助大数据时代着重打造具有档案特色的新媒体产品,在现有的档案网站和公众号建设的基础上,在众多新媒体平台开设官方账号,实现不同平台的整体联动,在多个平台上实现档案建设资源的统一共享,打造具有地方特色的档案产品宣传口,实现档案建设工作的与时俱进,打造良好的档案品牌形象。

### 3.2 档案部门自身建设

#### 3.2.1 馆藏档案数字化

馆藏档案数字化是档案馆适应智能数字化时代要求、建立数字化档案馆的必由之路,为了保证数字资源的共享和统一利用,需要对馆藏档案进行数字化处理。为了提高工作效率,可以着重对利用率较高、需求量大、具有长远使用价值的档案优先进行数字化处理,注意在数字化过程中需要进行全程监管,防止档案的完整性受到破坏,确保档案信息安全。

#### 3.2.2 提高档案工作人员素质

档案部门应当定期对档案工作人员展开档案编研工作的专门培训,提高工作人员的档案意识和业务能力水平,同时在全国范围内进行学习调研,总结成功经验,与相关高校开展合作,积极输送档案人才,实现档案工作队伍能力水平的提升。

### 3.3 技术保障

#### 3.3.1 数字编研技术

传统的档案编研工作位于档案管理工作的后端,内向性和滞后性较强,导致编研内容缺乏灵活的联系,无法满足前端工作的利用需求。档案的数据化编研技术从选材范围开始,将业务数据和数据化档案纳入到选材范围之中,以先进的人工智能技术开展编研工作,通过机器对数据库进行实时动态的更新,进一步扩大了档案数字编研产品的传播范围,与用户之间实现了良好的双向

在线互动,以多样化、多维化、动态化、开放化的数字编研产品为用户提供服务,向着智能化档案知识服务的方向前进。

#### 3.3.2 数据挖掘技术

在数据化档案管理的时代下,档案编研工作面临着更加标准化、结构化、开放化、支持更多智能机器操作的数据。编研工作需要利用更先进的数据挖掘技术对数据进行采集、清洗、挖掘处理,利用神经网络法、决策树法、遗传算法、粗糙集法、模糊集法、关联规则法等算法对数据进行处理,从传统的人工化处理向着机器自动化处理的数据化方向转型。<sup>[3]</sup>

#### 3.3.3 数字人文技术

将资源数字化、数据管理技术、可视化技术、交互技术引入到档案工作中,实现计算机技术与档案编研工作更加深入的融合,从平台媒介的传播转变为多维数字媒介传播;从文书档案编研转变为多元化数字编研,进一步实现转型升级以适应数字化时代的需求。

#### 3.3.4 信息安全保障条件

档案部门需要利用内外网隔离技术、防火墙技术、电子文件的数字签名、基于HTTP的加密传输技术、用户授权管理模式、密钥加密存储技术等多种途径给予用户数据全方位全过程的安全保护,防止用户个人信息泄漏,侵犯用户个人隐私,造成用户利益受损,保证用户信息收集工作的合法化开展。

### 结束语

档案是一种社会记忆,档案建设工作其本质上是围绕用户开展的,用户的需求分析应当是档案数字资源建设进行主题定位要考虑的首要因素。而大数据时代的到来使得数据信息数量急速膨胀,表现形式越来越多样,更新速度越来越迅速,互联网背景下的档案用户信息需求也越来越复杂,从而导致档案数字资源建设工作对于用户需求的分析工作提出了更高的要求,面对越来越复杂的用户信息和行为数据,需要利用更加先进的信息技术和更完善的模型流程来对用户信息进行采集挖掘、过滤筛选,最后进行分析处理形成知识图谱,为档案数字资源建设工作的精准定位提供方向指导。

### 参考文献

- [1] 徐爽. 大数据时代预测性报道的发展策略[D].安徽大学,2015.
- [2] 王芳. 大数据背景下数字档案资源建设研究[J]. 数字传媒研究, 2023, 40 (03): 14-17.
- [3] 崔艳红. 试论小数据在档案信息个性化服务中的应用[J]. 办公室业务, 2020(18): 86-87