

1. 大数据环境下档案信息资源的整合

随着互联网的普及,人们步入了信息化时代,RFID 物联网技术、web 文本挖掘技术、云计算等信息技术的高速发展,信息呈现爆炸式的增长,大数据随之而来,得到了各行各业的高度重视,档案行业也不例外。

而近些年来,随着社会信息化的飞速发展,人们对档案信息资源的需求也不断增长,大数据时代的来临也使档案信息资源发生了翻天覆地的变化。但是档案信息资源难以避免地受到馆藏类别以及地域的制约,已经无法适应与满足信息时代公众对档案信息资源的需求。所以,在一体化信息资源管理系统中纳入档案信息化建设,将封闭而又单一的档案信息资源转化成类别丰富、综合开放的档案信息,实现档案信息化以及档案信息资源共享显得尤其重要。因此,实现档案信息资源的挖掘与整合就显得尤其重要。而实现档案信息资源的整合,就有必要在“大档案理念”的指导下构建“大档案格局”以及档案信息资源多元化整合体系。

肖芃指出:“所谓大档案理念就是指档案信息资源是国家信息资源体系的一个组成部分,各地区按照各自的实际情况,将一个地区内分散保存在不同档案保管机构的档案资源集中统一到综合档案馆,实现档案资源的集中与整合,其基本理念就是‘多档合一’,即档案实体的集中以及档案资源和信息的集聚。借助这一理念,构建区域性大档案格局,也就是要打破档案保管机构之间有关档案接收利用的界限,广泛整合本区域范围内的全部档案资源,以方便各方面的使用。”^①

1.1 大数据环境下档案信息资源整合的必要性

我们将从信息化时代的发展趋势和提升档案信息资源服务效能的需求两方

^① 肖芃,林忠华,卜鉴民.构建区域性大档案格局[J].中国档案,2011,03:22-23.

面来进行必要性分析。

1.1.1 信息化时代的发展趋势

随着社会信息化的发展,数字化与网络化建设的不断完善,档案信息资源的记录载体、记录方式、管理方式也随着时代的进步而发生着变化,档案信息资源的管理也应该朝着网络化、数字化的方向发展。

随着人类的进步和发展,大数据时代的来临,人们在计算机系统存储的数据信息越来越多,这些数据是人们工作、生活和生产活动等的原始记录,能够为人们提供重要的利用价值。例如,美国沃尔玛超市将尿布与啤酒这两种看似毫无任何关联的商品摆放在一起进行销售,这一举措带来了意想不到的收益,使得超市尿布和啤酒的销量大幅增加。原来,美国的妇女通常在家带孩子,所以她们会经常嘱咐丈夫在下班路上为孩子买尿布,而丈夫在买尿布的同时就顺便购买了自己爱喝的啤酒。于是,沃尔玛就通过这种发现为企业带来了丰厚的利润。这个故事讲述了沃尔玛超市通过对自己企业的档案信息资源的数据信息进行挖掘,为超市的发展带来了黄金价值。因此,档案信息资源整合将是挖掘档案信息资源潜在信息价值的有效措施,是实现档案信息资源共享化的必然选择,也是适应社会信息化进程的需要,更是档案事业发展的必然趋势。

除此以外,实现档案信息资源的整合是解决传统档案资源管理模式带来的弊端的需要。长期以来,档案保管机构各自为政,造成档案资源长期分散,而这种分散性已然不适应大数据时代集中性的需求,于是便产生了对档案信息资源进行整合的诉求。档案信息资源数字化、信息化后,体现的明显特征是相对完整性、集中性,这就出现新的诉求——档案信息资源整合。尤其是现代电子计算机普遍应用,所生成的文件档案信息越来越具有电子特征,我们进行整合时不得不考虑到未来发展趋势问题。例如,科技部、财政部、农业部等有关部门协调成立的国家科技文献资源网络服务系统,教育部主持推进的全国高校信息保障系统,由文化部国家图书馆牵头的中国数字图书馆工程以及各地数字档案馆的建设等,这些工程建设是对信息完整性、集中性需求的体现。

1.1.2 提升档案信息资源服务效能的需求

国家档案局在2010年对《各级国家档案馆收集档案范围的规定》进行了修改,按照以人为本的思想,引导全国各级档案部门以民生需求为导向,把涉及人的档案应收尽收,建立面向全民的多元化档案资源体系。

在现代政府以公民需求为导向信息管理核心下,充分利用信息技术提供高效、高质档案信息服务,是未来服务发展的方向。在这种背景下,档案馆被推

向了信息公开的前台,意味着档案信息资源开发具有了政治合法性和迫切性。社会信息资源整合程度的提高与公众信息意识的觉醒为档案信息资源的整合创造了良好的社会环境与氛围,同时也使档案资源的整合成为了一种必然趋势。

近年涉及老百姓切身利益的民生档案数量与日俱增,与之相对应的是人民群众利用档案的需求也不断增加,因而迫切需要一种能够集中保管和统一利用的档案管理机制的出现与创新。整合档案信息资源为公众提供了一个双向主动式档案信息服务手段。除此以外,一方面是档案信息资源提供服务的频次、速度、要求越来越高;另一方面是档案信息资源服务的范围、空间、形式越来越广,社会的需求永远是激活档案信息资源整合和开发的力量源泉,推动档案信息资源整合的动力是适应时代发展和档案信息资源服务对象多元化的需要,档案信息资源整合的建设就会使档案服务社会的力度、方式、手段实现新跨越。

总之,实现档案信息资源的整合,可以提高人们利用档案信息资源的检索效率,可以改善档案网站、档案馆以及档案室等的服务质量。

1.2 大数据环境下档案信息资源整合的 SWOT 分析

大数据是以容量大、类型多、存取速度快、应用价值高为主要特征的数据集合,正快速发展为对数量巨大、来源分散、格式多样的数据进行采集、存储和关联分析,从中发现新知识、创造新价值、提升新能力的新一代信息技术和服务业态。^①

随着互联网的普及,计算机信息技术和网络通讯技术的飞跃式发展,各种数据和信息呈现出爆发式的增长。事物都有两面性,互联网在带给人们获取大量文本信息资源快捷方便的同时,也带来了一些难题,如如何快速有效地在海量的信息资源中挖掘出自己所需要的信息资源。总之,大数据时代已经悄然降临,海量信息也给档案部门的档案信息资源整合带来了挑战。因此,档案部门应该实事求是地立足于档案信息资源整合的现状,结合大数据的时代背景,充分应用大数据时代的信息挖掘技术,采取有效的措施实现档案信息资源的整合。

接下来,我们将采用 SWOT 分析法:S(strengths)是优势;W(weaknesses)是劣势;O(opportunities)是机会;T(threats)是威胁或挑

^① 国务院. 国务院关于印发促进大数据发展行动纲要的通知 [EB/OL]. <http://www.gov.cn/zhengce/content/2015-09/05/content-10137.htm>

战。对大数据环境下档案信息资源整合的优势、劣势、面临的机遇和挑战进行分析，如图 1.1 所示。

优势 (strengths)	劣势 (weaknesses)
(1) 国家政策的支持 (2) 信息挖掘技术的进步 (3) 数字化信息管理系统的发展	(1) 缺乏统一的技术整合规范标准 (2) 理念与时间进度不匹配 (3) 档案部门各自为政，缺乏有效沟通联系
(1) 信息化建设 (2) 电子政务建设 (3) 数字化建设	(1) 资金缺乏 (2) 档案信息资源质量难把关 (3) 体制建立困难
机遇 (opportunities)	挑战 (threats)

图 1.1 SWOT 分析矩阵

1.2.1 优势分析

在大数据时代，各种数据信息日益膨胀，呈现爆炸式的增长。那么，将各种分散的、独立的信息统一到一起，形成逻辑上统一的整合体，将是目前档案部门存储和利用大数据的最大需求。因此，实现档案信息资源的整合与共享是大数据时代和信息化时代的召唤，得到了国家层面的高度重视。接下来，我们将从国家政策支持、信息挖掘技术进步、数字化现代管理系统的产生三方面来分析大数据时代档案信息资源整合的优势。

1.2.1.1 国家政策的支持

国家档案局在 2010 年对《各级国家档案馆收集档案范围的规定》进行了修改，按照以人为本的思想，引导全国各级档案部门以民生需求为导向，把涉及人的档案应收尽收，建立面向全民的多元化档案资源体系。由此，可以看到国家是支持对档案信息资源进行整合的。2015 年 8 月 19 日，国务院审议通过了《促进大数据发展行动纲要》，这标志着我国在顶层设计上对大数据的实践与实施作出了集体部署。《促进大数据发展行动纲要》不仅对其背景及意义进

行了阐述,对涉及的主要内容及产业也进行了部署。其中,虽然没有明确档案部门的任务,但指出如何应对大数据时代下的快速变革是档案部门的当下之责^①。

“十三五”规划期间,国家档案局印发的《全国档案事业发展“十三五”规划纲要》中也制订了相关的政策支持档案信息资源整合的发展。其中,有一项发展目标就是“实现档案资源多样化。依法管理档案资源,各级国家机关、团体、企业事业单位档案实现应归尽归、应收尽收;档案资源更加齐全完整、丰富多元,覆盖人民群众的档案资源体系更加完善”^②。

此外,地方也主张区域性“大档案格局”,颁布了相应的政策。以苏州为例,从2010年10月1日起正式实施的苏州地方性档案法规《苏州档案条例》中就明确规定,各级开发区、乡镇(街道)也可以建立档案馆,集中统一管理本区域范围内所有档案。《苏州档案条例》第七条规定“有条件的镇(街道)设立档案馆”,第八条则明确指出“国家级、省级开发区管理机构应当建立档案馆,负责本单位的档案工作,并对所属单位的档案工作进行监督和指导。”管理模式和管理方法的转变,必将推动开发区和乡镇(街道)的档案工作进一步向法制化、规范化迈进,从法规层面赋予了构建区域性大档案格局实践的合法性。^③

1.2.1.2 大数据环境下信息挖掘技术的进步

互联网的发展与普及,各种数据以及信息呈现出爆发式的增长。互联网在给人们获取大量文本信息资源快捷方便的同时,也带来了诸如如何挖掘、筛选自己所需信息的难题,但是大数据时代的信息挖掘技术则刚好可以帮助人们解决这一难题。以“web”数据挖掘技术为例。

首先,谈一下“web文本挖掘技术”在档案信息资源整合中的应用。Web是当今互联网上最受欢迎、最为流行的超文本信息系统,不仅能实现各种类型数据的无缝集成,还具有提供图形界面快速检索等功能。因此,我们可以利用web文本挖掘技术的高效率、智能化等优势,结合档案信息资源整合中所面临的多种多样的问题,帮助人们改善检索效果以及服务。

接下来,介绍一下文本挖掘的基本思想,如图1.2所示。

① 许桂清.《促进大数据发展行动纲要》助力档案工作[J].中国档案,2015(11):20-21.

② 国务院.国务院关于印发促进大数据发展行动纲要的通知[EB/OL].<http://www.gov.cn/zhengce/content/2015-09/05/content-10137.htm>

③ 肖芃,林忠华,卜鉴民.构建区域性大档案格局[J].中国档案,2011,03:22-23.

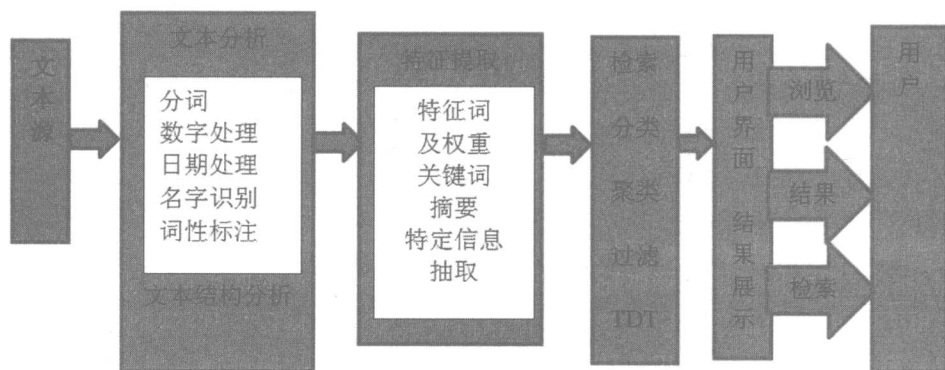


图 1.2 文本挖掘的基本思想

一方面，web 文本挖掘技术可以有效地改善档案信息资源的效果。通过对信息的聚类处理，对使用者的检索历史信息进行分析，对分析结果进行分组，并分别进行标注，可以使提供的检索内容更加精确，一定程度上优化使用者的检索效果。这样可以提高检索信息的准确度，提升搜索效率，节约检索时间等。另一方面，可以改善服务。如档案馆、档案网站等可以通过掌握、分析使用者浏览各类信息资源的频率以及所花费的时间，就可以分析判断出使用者对不同类型信息资源的兴趣度，有利于升级完善更加人性化的推荐、定制等多样性的服务。

1.2.1.3 数字化的现代档案管理系统的产生

随着计算机技术、通信技术以及互联网技术的飞速发展，传统的档案管理模式也遇到了严峻的挑战。与发达国家相比，我国档案管理现代化建设还存在着滞后性。在 2006 年，有关组织部门明确指出：要充分利用现代技术改造传统的档案管理方式，加快电子档案建设，完善干部档案管理系统和干部信息管理系统，逐步实现档案管理的数字化。档案数字化必将成为今后档案的主要存在形式。

数字化档案管理是对传统干部人事档案管理工作的一次创新，能够实现对档案和档案材料收集、鉴别、整理、保管、转递、统计、查阅等日常工作的数字化管理，并可通过组织系统专网实现干部档案的网上浏览和远程查询。这为今后干部任用、干部提拔等工作带来了极大的方便。系统分为日常业务管理、档案数字化采集系统和数字化档案查阅系统三大部分。

以苏州构建区域性大档案格局为例，苏州市的现代化档案管理系统已经基本建成。肖芄在文章中提到，“苏州市具有较完备的网络基础设施，各机关单位统一管理系统和硬件设施，电子政务的运行已经普及，公文从网上流转、传

输、归档,实施网上无纸化办公;苏州市数字档案馆项目已经完成,能通过网络处理档案馆的所有业务,而各机关事业单位的信息化建设也已初见成效,不少单位档案已经全文数字化处理,建成数字档案室,形成了统一的数字档案管理标准,建立了一批档案工作的优秀管理人才,为构建区域性大档案格局打下了良好技术基础”^①。

1.2.2 劣势分析

凡事都有两面性,有利就有弊,接下来,我们将从三个方面集中分析大数据环境下档案信息资源整合的劣势。

1.2.2.1 缺乏统一的技术整合规范标准

许多学者都认为技术与标准是制约档案信息资源整合的重要“瓶颈”。熊志云认为,“尽管我国档案信息化建设起步较晚,但对档案信息的标准化工作应该是一开始就比较重视的”^②。梁作华指出,“迄今为止,标准问题仍未受到足够重视,缺乏制定标准的统一机构,国家尚未出台档案数据库结构、信息存贮和著录格式、软硬件配置、网络体系结构、信息处理界面等方面标准,要建立一个比较完整的档案信息资源整合的标准体系,还需一个长期的探索和实践过程”^③。许多研究者都指出,在档案信息化建设中,标准问题仍然没有受到足够的重视,缺乏一个统一的机构对此进行统一的研究与组织,也就谈不上建立跨系统、跨领域的标准制定机构了。因此,建立统一专门的机构,在全国范围内逐步推出和建立比较完善的信息资源管理标准体系,已成为档案信息资源以及其他信息资源整合工作的当务之急。

1.2.2.2 理念与实践进度不匹配

“大档案”理念由来已久,但是在实践进程上并未得到良好的落实。长期以来,许多档案工作人员和专家学者都提到对档案信息资源整合,应该转变档案管理观念,但是如何转变并没有一套详细具体的对策。因此,我们不仅要提出与时代发展相匹配的管理理念,还要将理念贯彻到档案信息资源整合的实践之中去。

在大数据时代,档案信息资源整合需要各个环节的配合和各个部门协调一致的工作才能完成,各部门、各环节的协调都需要一个强大的信息系统来组

① 肖芃,林忠华,卜鉴民.构建区域性大档案格局[J].中国档案,2011,03:22-23.

② 熊志云.档案信息资源的整合趋势及整合措施浅议[J].档案学研究,2005(1):29-33.

③ 梁作华.浅谈网络环境下档案信息资源整合策略[J].山东省青年管理干部学院学报,2008(6):156-157.

织、协调和实现信息资源共享,使得档案信息资源得到最优化的利用。使得良好的档案信息资源整合观念,发挥意识的能动作用,更好地指导档案信息资源整合的实践工作。

1.2.2.3 档案部门各自为政,缺乏有效沟通联系

目前,中国实行的档案管理体制是条块相结合的,即综合档案馆主要负责党和政府机构形成的档案,而专业档案则由专业档案机构负责保管,档案部门由同级档案行政管理部门进行指导、检查和监督。无论是在实践层面还是在法规层面,条块分割已经成为档案管理的基本形态,并且已经持续多年。但随着档案信息资源整合工作的开展,许多学者都指出这一制度阻碍了档案信息的共享。郑鸥认为“它虽然有其合理性,但也造成了地方各级国家综合档案馆在档案资源整合上难以有所作为。由于档案资源的分散、馆藏档案信息的‘含金量’不高,很难满足日益增长的为领导决策资政服务、为群众维权服务的需求”^①。

1.2.3 机遇分析

只有抓住机遇,才能更好地发展自己,大数据环境下档案信息资源整合也面临着一些机遇,抓住这些机遇,有利于档案事业更好的发展。接下来,将从三方面进行机遇分析。

1.2.3.1 我国的信息化建设为档案信息资源提供了环境支撑

我国进行信息化建设中通信网络与计算机不断完善为档案信息资源的整合提供了良好的基础设施条件。首先,我国已经基本上形成了一个覆盖全国、通达世界、技术先进、业务多样的国家公用通信网络框架。目前,基本全国所有的学校、科研部门、政府、企业和家庭都普遍使用计算机网络。其次,网络的连通是档案资源整合的基本手段和有效途径,通过网络使数字化的档案资源互相联结,社会公众能够不受时空条件的约束而获知和获取自己所需的档案信息。先进专业的通信网络的建成为档案信息资源整合提供了许多实践的机会,日渐成熟的因特网技术为建设档案信息整合提供良好的基础设施条件。

除此以外,“十二五”期间,我国的档案信息化建设已经取得了一些成果,例如,“档案信息化建设初具规模。初步建成以局域网、政务网、因特网为平台,以档案信息管理系统为支撑,以档案目录中心、基础数据库、档案利用平台、档案网站信息发布为基础的档案信息化体系。”

^① 郑鸥. 加强档案信息资源整合势在必行 [J]. 新上海档案, 2006 (05): 22-25.

“十三五”规划中,国家档案局印发的《全国档案事业发展“十三五”规划纲要》中也提到了要加快档案管理信息化进程的目标。它要求各档案部门、档案网站等主体“加快提升电子档案管理水平,积极参与国家政务信息化工程建设,制订相关标准和规范,明确各类办公系统、业务系统产生的电子文件归档范围和电子档案的构成要求;加强对业务系统电子文件归档管理,通过推进电子会计档案管理促进电子政务和电子商务文件归档管理工作”。

1.2.3.2 电子政务建设为档案信息资源整合提供了技术支持

胡锦涛总书记在党的十七大报告中提出“加快行政管理体制改革,建设服务型政府”,并明确指出要“健全政府职责体系,完善公共服务体系,推行电子政务,强化社会管理和公共服务”。实施电子政务是改革开放的需要,是建设服务型政府的基础工作。除此以外,《国家电子政务工程建设项目档案管理暂行办法》也具体规范了档案管理中相应的电子政务建设的事项。

在我国,档案有着为政治提供服务的职能,因而档案馆、档案室等档案管理部门也必须结合国家电子政务建设的大背景去开展档案管理工作。因为电子政务建设能够促使政府更加重视档案室、档案馆、档案网站、档案公共服务平台等的建设,给予相应的资金以及技术支持。有了政府资金、技术支持作为后盾,更加有利于档案信息资源整合工作的顺利开展。

1.2.3.3 档案部门的数字化建设为档案信息资源整合提供了基础

近年来,社会环境的变革使档案界不得不对档案信息工作进行改善,各地档案部门不同程度地进行了数字化、信息化管理建设,为档案信息资源整合提供好的开端。档案部门要努力建设并投入使用一批内部局域网,基本实现档案管理现代化和办公自动化,部门、地方自行建立为本部门机关服务的档案目录信息中心,为逐步构建中国档案文献数据库创造条件。档案部门根据电子档案管理的要求,加强对本单位电子文件环节管理,保证电子文件真实、完整、有效。此外,北京、上海、江苏、浙江、广东等地选择在国家档案馆开展网络环境下接收电子档案试点工作。资料显示,北京市近年来制订档案整合的发展规划,建设市区县档案馆之间的联网与档案目录开放;深圳市建立数字档案馆系统,将原有各种载体的档案数字化,对档案文件实施数字化管理。

1.2.4 挑战分析

知己知彼,方能百战不殆。一方面,我们既要抓住机遇,促进档案信息资源整合开展;另一方面,也要了解其面临的挑战和威胁,以制订好相应的解决对策。接下来,我们将从以下几方面进行分析。

1.2.4.1 需要庞大的资金支持

虽然,档案管理部门在进行档案信息资源整合工作的时候,能够得到政府的财政资金支持,但是这是远远不够的,需要更多的资金为档案信息资源整合的后续工作做支撑。而档案部门也缺乏其他的筹款渠道,因而资金将是大数据环境下档案信息资源整合面临的一大困境。此外,受我国各地区经济发展水平的制约,各区域档案信息资源整合建设的投入、档案数字化、信息化建设等方面都存在着严重的不平衡、信息不对称等问题。

与此同时,滕霞认为,“档案信息资源整合还有的存在论证不足问题,在系统设计档案信息资源整合规划时没有一个相对完整的构想,对整合过程中容易出现的问题或解决对此准备不充分,论证存在缺陷,所以其结果既达不到整合的要求又造成不必要的浪费”^①。

1.2.4.2 档案信息资源质量的把关

档案信息资源的质量难以控制。蒋冠指出,“就目前的情况来看,我国大部分公共档案馆的馆藏结构单一,政府部门的文书档案占了很大的比例,而其他类型的档案少之又少,特别是某些特色档案更是缺乏,档案信息库的建设要有坚实的资源基础,单调无用的信息根本无法吸引用户的目光,大量的投入只能是一种浪费。因此,要采取多种措施,大力加强档案馆的馆藏建设,丰富自身资源,建设特色档案库已成为了目前工作中的重中之重。另一方面,档案信息资源与其他社会信息资源相比,有着机密性的特点,有一大部分档案因为涉及国家集体的秘密而不能轻易地在网上公布,而必须履行一定的手续,超过一定的时限才可以公之于众,而事实上,等到这些档案过了保密期之后,其所载信息则往往已经过时滞后,无法满足公众的现时需要”^②。除此以外,档案信息资源的质量没有统一而具体的标准。所以,在进行档案信息资源整合的时候,如何收集质量优、价值高的档案资源也将是面临的一大挑战。

1.2.4.3 档案信息资源整合管理体制的建立存在困难

我国现存档案工作实行的分级、分专业管理体制,暴露许多弊端,已经不适应档案信息资源整合工作的开展,大量档案资源长期散存在各单位,处于非专业管理状况,管理混乱,损失难以想象。有的单位把档案长期堆放在楼层间,虫蛀鼠咬,温湿无常;有的单位档案就堆放在各个办公室,基本无人过问,损毁现象严重。由于档案管理无序,查找利用困难,更难以服务社会,并存在重复建设的问题,造成财力、物力浪费严重,同时,综合档案馆馆藏结构

① 滕霞. 档案信息资源整合对策探析 [D]. 哈尔滨: 黑龙江大学, 2008: 22-23.

② 蒋冠. 网络环境下档案信息资源整合研究 [D]. 湘潭: 湘潭大学, 2005: 28-29.

单一,服务功能弱化。^①

实施整合档案管理机制不仅是对已有的档案信息资源的整合,重要的是通过资源整合理顺管理体制和建立起新的规范化运行机制,为确保以后不再造成新的档案资源分散,建议实行大档案管理体制,确定省、市级综合档案馆的中心地位,并赋予其整合的权利,把现有市县级综合档案馆作为中心基点,才能有效地克服资源分散的问题、打破各自为政的管理模式,提高档案馆的辐射力、影响力、凝聚力。

1.3 大数据环境下档案信息资源整合的措施

我们将从档案实体整合与数字档案信息资源整合两方面来提出应对措施。首先,来谈一下档案实体的整合。

1.3.1 档案实体整合

档案实体整合是一个个体层次的整合过程,丰富的馆藏是档案信息资源整合的基础。档案实体整合包括综合档案馆整合自身管理制度、管理程序、馆藏系统的信息,还要包括县级区域内各种实体信息部门的整合,将区域内各个独立、分散的部门档案源进行综合整合。

1.3.1.1 现有馆藏整合

档案馆不再仅仅是一个实体保管机构,还是今后实现档案资源共享的主要源头和基地,一般传统的档案实体以案卷形式保管在库房,应对其馆藏数据清楚掌握,做好基础的编目工作。目前,档案馆的实体整理工作一直在做,但是结果不尽如人意。根据档案整合功能特征从档案馆管理制度化、归档程序化、馆藏数字化、利用网络化、控制智能化方面进行管理。做好现有馆藏各种载体标准、海量存储整合工作,有选择地将原始馆藏中有特色、有较高利用价值的档案数字化,积极把已接收进馆的文件建立编研成果数据库,使传统档案信息与现有的档案信息共同发挥作用,如电子政务档案、城建档案、指纹档案、民生档案等特色数据库建立。

1.3.1.2 开展区域档案信息资源整合

确立档案形成的一个县级的部门为区域展开。以往是单个部门保存自己形成的档案,造成分散性单个保存的条件有限,如果把一个区域县级的部门档案

^① 滕霞. 档案信息资源整合对策探寻 [D]. 哈尔滨: 黑龙江大学, 2008, 21.

整合在一起会节约很多人力、物力。滕霞认为,“在区域整体规划中设立县级单位为档案管理中心,各级档案信息形成部门向县档案馆移交,建立一个以档案部门为主体、各专业主管部门配合的区域管理模式,实现档案资源集约化、人员素质现代化、业务建设标准化、管理工作规范化、利用服务优质化”^①。

1.3.2 数字档案信息资源整合

在大数据的时代背景下,档案数字资源具有数量庞大、增长迅速、多源异构等新特点,在给人们带来丰富的信息的同时,也给数字档案信息资源的整合带来了一定的困难,如数据存储问题、安全保障体系的缺失等问题。接下来,我们将从以下几方面对大数据时代数字档案信息资源的整合策略进行探讨。

1.3.2.1 实现由馆藏中心模式向服务中心模式的转变

上述我们提到过大数据时代的信息挖掘技术,如云计算、web 2.0 文本挖掘技术等。这些大数据技术可以通过对复杂关联的数据网络中出现的趋势进行预测,从而为人们的行为决策提供有益指导。这就要求档案部门要改变过去单一的“供给式”的思维模式,关注大众的利用需求,构建起以社会利用需求为导向的档案数字资源体系。如档案网站导航、索引等人性化服务的提升都可以更加方便用户并时刻关注用户需求的变化,实现由馆藏中心模式向服务中心模式转变,不断提高档案服务与用户之间的匹配度。

1.3.2.2 构建适应大数据要求的档案数字资源分析系统

毫无疑问,构建适应大数据要求的档案数字资源分析系统依然要用到大数据时代的信息挖掘技术。接下来以“云计算”技术为例来说明。

云计算技术具有虚拟资源化、高可扩展性、高可靠性、按需付费、泛在接入等显著的特征。它适应了大数据时代分布式存储与海量数据并行处理的需要,实现了计算机资源的服务化,是大数据时代档案数字资源整合的基础平台和支撑技术。

首先,各档案部门应根据国家统一规划以及自身基础设施建设与档案数据库资源匹配程度的具体情况,灵活选择适合的云部署方案。对于那些处在档案数字资源整合关键节点的部门应架设私有云,其他部门可根据自身情况将关键数据存放在私有云,同时,以动态申请公有云的方式弥补自身计算能力、存储空间等的不足。其次,云计算能统一各应用环境之间的业务逻辑、组织结构和表达方式等,消除信息孤岛,从而建立集成的档案数字资源管理平台,促成档

^① 滕霞. 档案信息资源整合对策探寻 [D]. 哈尔滨: 黑龙江大学, 2008. 46—47.

案数字资源深层次整合与知识开发的实现。再次,云计算能实现对档案应用的整合,并以服务的形式向用户发布;同时,支持用户利用各种终端设备随时随地访问所需的云服务。这些都将在最大程度上发挥档案数字资源整合的优势,提高档案服务的效率和便捷程度。

1.3.2.3 强大数据时代档案数字资源整合的安全保障体系建设

首先,应建立 IAM(身份和访问管理)和隐私保护系统,实现统一身份认证与访问权限控制,达到用户安全集成管理的目标,有效应对档案数字资源整合与大数据应用过程中的安全风险。其次,通过数据加密技术保护档案信息安全。通过 SSL(Secure Sockets Layer,安全套接层协议层)加密,实现在数据集的节点和应用程序之间移动保护大数据。^①再次,综合运用大数据技术手段与安全保密制度,加强对重点领域档案数据的日常监管,有效应对档案数据聚集性与档案利用需求无序性造成的档案泄密风险。最后,实时开展档案数字资源异地异质备份工作,提高系统容灾能力。

1.3.2.4 建立统一的档案数字信息资源整合标准体系

在大数据时代,档案数据的多样性已成为常态,要实现档案数字资源的整合就需要协调相关利益方建立起兼顾适用性、稳定性和国际性的档案数字资源整合的标准体系,完成对不同协议、标准、规范的整合。这包括档案信息化过程中涉及的各类数据组织方式和网络通信协议的整合,各相关业务系统中使用的数据标准和协议规范的整合,以及采用的各类存储、应用标准的整合等。唯有如此,才能确保整合工作遵循相同的标准,方便档案数字资源的存储和迁移,实现档案数字资源的交流与共享。

^① 郭三强,郭燕锦.大数据环境下的数据安全研究[J].科技广场,2013(02):28-31.

2. 大数据技术在档案信息资源挖掘中的应用

自 2012 年英国大数据学者迈尔·舍恩伯格在《大数据时代》中首先提出大数据概念后,这一概念迅速成为了当代的热门话题,如今大数据时代已经来临,大数据思维深刻改变着人们的思想、工作和生活方式,大数据技术深入社会的各个领域并发挥出巨大的作用。随着计算机技术、互联网技术的快速发展以及云计算技术的进步与普及,我们进入了大数据时代。大数据具备种类多、总量大、价值丰富和变化迅速等特点,随着大数据时代的到来,各类信息资源也渐渐具备大数据的特点,包括档案信息资源。使用传统方法对大数据化的档案信息资源进行挖掘开发,从而实现高效利用难度渐增,因而必须考虑将大数据技术应用于该领域,以实现档案信息资源挖掘工作的进步。

在大数据时代,伴随着互联网技术的飞速发展,各类社会媒体的普遍应用,数据呈现出数量大、种类多、变化迅速、价值总量高的特点,档案信息资源作为最重要的社会信息资源之一,也呈现出大数据特点。在大数据背景下,档案信息资源的数量急剧增长,种类愈发繁杂,数字化、信息化程度不断提升,使用传统的管理手段已经难以处理新形态的档案信息资源,树立起大数据观念下的档案信息资源挖掘新思维、构建大数据技术指导下的档案信息资源挖掘新技术显得愈发必要。

目前云计算、语义处理技术和可视化技术等新型大数据应用技术已经趋于成熟,并在档案管理方面得到了实践操作。深入探究这些技术在档案信息资源管理中的应用,总结经验和不足,有利于档案工作更好地发展。

以大数据技术指导档案信息资源挖掘工作,对档案信息资源的价值进行深入分析和系统全面地提取,对于未来档案服务工作具有重大意义。

2.1 大数据技术在档案领域的应用背景

大数据时代数据的种类和规模都空前庞大,成为了一种最重要的社会资源,且亟待人们对其进行开发和利用。大数据时代深入改变了人们的生活、生

产和思维方式,对社会各方面造成了巨大影响,档案信息资源在新的社会背景下也发生了巨大改变并愈发显现出大数据的特征,如何对海量档案信息资源进行高效系统挖掘,从而实现深层次开发利用成为当下档案工作的中心。传统的档案信息资源挖掘工作不能满足新形势下档案信息资源的开发要求,将以云计算、语义引擎和可视化分析为代表的大数据技术应用到档案信息资源的挖掘工作中,可以为其带来巨大机遇,世界各国对于大数据技术深入推广、积极倡导,我国也出台相关政策进行支持,为大数据技术深入应用在档案信息资源挖掘领域提供了支持。

2.1.1 大数据技术为档案信息资源挖掘工作带来新机遇

国际咨询机构麦肯锡对大数据做出以下定义:“大数据是指无法在一定时间内用传统数据库软件工具对其内容进行采集、存储、管理和分析的数据集合^①。”因此,在大数据时代必须要使用新的数据处理技术才能实现对数据资源更好地开发和利用。大数据背景下档案信息资源也具备了大数据的特征,主要体现在以下三点:一是各级档案机构所产生的档案信息资源总量日渐庞大且增长迅速;二是档案信息资源种类日趋繁杂,而且结构日渐复杂;三是档案信息资源的价值丰裕度、凝聚度很高^②。对具备大数据特征的海量档案信息资源进行广泛采集,深入挖掘,对档案信息资源发挥最大化效用具有不可估量的意义。

档案信息资源的挖掘工作是指对海量的档案信息资源进行采集,并对采集到的数据进行清洗、集成、变换等处理,最后选择相应的挖掘模型,实现对档案信息资源价值的开发和提取,从大量的档案信息资源中挖掘价值、提取知识,从而实现其更为广泛和高效地利用的过程。档案信息资源的挖掘流程如图2.1所示^③。

① 石峻峰,周俐霞,等.大数据时代高校数字档案信息资源管理研究[J].现代教育技术,2015(1):19.

② 周枫.资源·技术·思维:大数据时代档案馆的三维解释[J].档案学研究,2013(6):61.

③ 王静,辛玉明,高鸿雁.档案数据挖掘中数据采集与准备问题浅析[J].现代情报,2016(6):73.

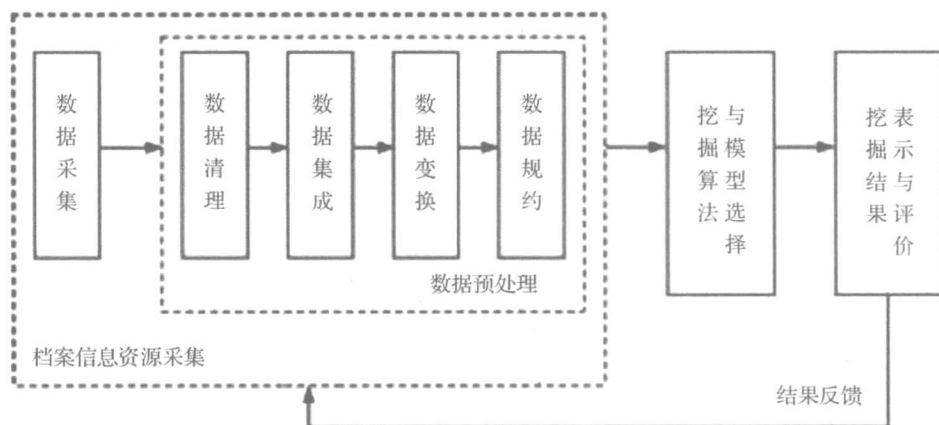


图 2.1 档案信息资源挖掘流程

档案信息资源的大数据化给其挖掘工作带来了很大困难，如档案信息资源的采集问题、清洗问题、价值分析问题和结果提取问题等，但是大数据技术的使用也给档案信息资源的挖掘工作带来巨大机遇，主要体现在以下三点：

1. 大数据技术可以实现档案信息资源更系统、全面的采集。大数据处理技术强调对整体数据进行分析 and 挖掘，以此取代传统档案信息挖掘中以抽样代替整体的方法，可以改变因为遵循传统经验思维搜集局部档案信息进行分析而造成的挖掘成果的片面性和不完整性^①。云存储技术手段为信息采集提供了足量的空间，为档案信息资源的系统、全面采集提供技术支持。

2. 大数据技术可以实现档案信息资源的智能化提取，并提高挖掘的精确度和效率。基于云计算的大数据价值分析技术可以在挖掘过程中提高精确度，可视化技术则对档案信息资源进行全面直观的呈现，语义处理技术为档案信息资源的智能检索创造了条件，有利于挖掘效率的提升。

3. 使用大数据技术对档案信息资源进行挖掘，可以弥补由于档案缺失而造成挖掘结果价值低的问题。大数据技术通过对海量档案信息资源进行处理分析，创建数据资源库，在某一部分档案信息资源存在缺失时，可以根据档案信息资源间的关联性原则对相关资源进行追踪，以补充缺失的档案信息，以保证档案信息资源挖掘结果的完整性和可靠性。

① 刘玉兰，韩海涛. 大数据时代档案工作面临的挑战与机遇 [J]. 兰台世界，2015（4）：15.

2.1.2 国家政策引领与支持

大数据概念自提出伊始,就成为最热门的名词之一。大数据技术给社会带来强烈冲击,深入影响着社会的各个领域并引发思想变革。2012年3月,美国政府发布“大数据研究发展倡议”,在大数据技术研究领域投资2亿美元,并将大数据上升到国家战略级别^①。2012年5月,联合国公布了名为“大数据的机遇与挑战”的白皮书,对于大数据技术给人类社会带来的机遇和挑战进行分析,在该报告中,联合国分析了大数据技术在中国互联网行业的发展状况,并认为大数据技术将会给中国互联网行业带来巨大的发展机遇。

2015年8月,我国国务院发布了《关于促进大数据发展行动纲要的通知》,在此通知中指出了我国大数据技术发展的形势和意义,认为大数据成为重塑国家竞争优势的新机遇,并提出了在我国发展大数据的指导思想和总体目标。这份通知提出在未来的国家发展过程中,应利用好我国的数据数量优势,努力实现数据数量、质量和数据应用水平的协同发展,注重对数据资源潜在价值的挖掘,将大数据这一战略资源的作用得到最大限度的发挥,以提升国家竞争力。

在这份《促进大数据发展行动纲要》中树立了未来发展大数据的指导思想,包括“大力推进政府信息系统和公共互联开放共享,消除信息孤岛,着力推进数据汇集和挖掘,推进数据资源向社会开放^②”,这些指导思想对于档案信息资源挖掘过程中使用以云计算为代表的大数据技术,实现档案信息资源共享、消除档案信息资源孤岛、实现数据广域采集都具有引导作用。

通知在总体目标中指出要在2017年之前实现跨部门的数据信息共享格局,扩大公共事业单位的服务质量和范围,完善教育、文化行业的大数据建设和管理。“完善教育、管理公众服务平台,推动教育基础数据伴随式收集和全国互通共享。”“建立终身电子学籍档案”“加强数字图书馆、数字档案馆……公益设施建设^③。”

目前,我国已经认识到大数据对于国家未来发展的重要价值,并为大数据技术的发展提供思想指导和政策支持。档案信息资源是国家记忆的主要构成部

① 刘智慧,张泉灵.大数据技术研究综述[J].浙江大学学报,2014(6):958.

② 中国政府网. <http://www.gov.cn/zhengce/content/2015-09/05/content-10137.htm> [2016-4-22 8:27]

③ 中国政府网. <http://www.gov.cn/zhengce/content/2015-09/05/content-10137.htm> [2016-4-22 8:28]