

The background of the cover is a deep blue color. At the top, there are several lines of binary code (0s and 1s) in a light blue, slightly faded font, arranged in a perspective that recedes into the distance. Below the binary code, there is a horizontal band with a filmstrip-like border. Inside this band, the title '数字档案馆建设与管理' is written in large, bold, yellow characters with a black outline. Below the title band, the author's name '李泽锋/著' is printed in white. The bottom half of the cover features a stack of overlapping CD or DVD discs, also in shades of blue, creating a sense of depth and digital storage.

数字档案馆建设与管理

◎李泽锋/著

中国档案出版社

责任编辑/罗 京

装帧设计/田奇丁

ISBN 7-80166-619-4



9 787801 666192 >

ISBN 7-80166-619-4

定价：16.00元

数字档案馆建设与管理

李泽锋 著

中国档案出版社

责任编辑/ 罗京
装帧设计/ 田奇丁

图书在版编目(CIP)数据

数字档案馆建设与管理/ 李泽锋著. - 北京:中国档案出版社,
2005.12

ISBN 7-80166-619-4

I. 数… II. 李… III. 数字技术 - 应用 - 档案馆
IV. G270.7

中国版本图书馆 CIP 数据核字(2005)第 161220 号

Shuzidangangan jianshe yu guanli

出版/ 中国档案出版社(北京市宣武区永安路 106 号 100050)

发行/ 中国档案出版社

印刷/ 河南地质彩色印刷厂

规格/ 890 × 1240 毫米 1/32 印张/8.25

版次/ 2005 年 12 月第 1 版 2005 年 12 月第 1 次印刷

字数/ 220 千字

定价/ 16.00 元

前 言

当前对电子文件进行收集、保管和保护，有效实现档案资源共享，将浩如烟海的珍贵档案进行全面数字化处理，高效便捷地为社会公众所利用成为档案馆迫切需要解决的问题。

随着信息化建设的迅速发展，档案馆的数字化建设需求逐渐提到日程上来。2000年12月，国家档案局、中央档案馆印发了《全国档案事业发展“十五”计划》，把“加快档案信息化建设”、“加快现有档案的数字化进程”列入其中。同月，国家档案局在中国档案管理现代化建设研讨会上宣布“十五”期间我国将加快数字化档案馆的建设。2002年，国家档案局作为成员单位加入国务院电子政务协调小组。同年，国家档案局制定了《全国档案信息化建设实施纲要》。《纲要》指出：在部分中心城市建设示范性数字档案馆是“十五”期间全国档案信息化建设的主要任务之一。

可以说，数字档案馆是未来档案馆的发展方向，这一点在档案界已经取得了共识。但数字档案馆到底应该是一种什么样的形态，在实际环境中如何建设，却是见仁见智。许多学者提出了富有创见的新思想、新观点，在许多方面进行了探索、实践。这充分说明了数字档案馆在档案学研究和实践中的活力。

本书是作者在完成了若干软件项目的开发以及所承担的郑州航空工业管理学院青年基金课题“企业档案馆数字化建设”后，将研究成果与平时发表的论文系统化的结果。因此本书内容在一定程度上反映了作者近年来对电子文件管理和数字档案馆这一问题的思考。

数字档案馆涉及面十分广泛，要研究的内容极其丰富，一些概

念和理论尚未定论。本书并不着力谈论这些理论性问题，只是把自己的观点提出来，而将研究的重点放在数字档案馆的建设实践方面，侧重于数字档案馆的基础设施建设与运行管理，以帮助正在研究与实践数字档案馆建设的实际工作者。同时，对于档案界研究不多的元数据与知识管理提出了自己的观点。作者试图在档案理论与信息技术之间寻找一个相对完美的结合点，争取对数字档案馆的实践提出一些容易操作的方法，以对档案学理论的进一步研究有所促进。

本书主要分为三个部分：第一部分为第一章，是对数字档案馆基本理论的分析，意在引导读者从整体上理解数字档案馆。第二部分以数字档案馆建设的顺序为线索，分别讨论网络平台建设、数字档案馆综合管理系统、档案网站建设与档案信息资源建设等四个方面，即第二、三、四、五共四章，使读者对于数字档案馆的基础设施建设有一个全面的理解。其中一些内容是作者在软件项目开发过程中的实践总结，旨在增强本书的实践性与可操作性。第三部分则是对数字档案馆管理的分析，设有“元数据管理”、“电子文件管理”、“信息安全”与“知识管理”章节，即本书的第六至九章。该部分出于开阔思路、扩大数字档案馆作用的目的，进一步将新的信息技术引入到数字档案馆的管理中，并提出了相应的解决方案。

由于数字档案馆是近年来才开始的较新研究课题，其研究的基本理论和相关内容尚未成熟，再加上信息技术的飞速发展，书中作者的一些见解会有一些偏颇之处，恳请档案界的专家学者与实践工作人员予以指正。本书在写作的过程中，广泛参考、借鉴了国内外学者的有关观点与成果。所引用的研究成果，尽可能地在参考文献中列出。

本书在写作过程中，刘国华教授、郝伟斌先生与李海涛先生提供了一些有益的建议，郑州航空工业管理学院 020462 班徐海秋等同学协助进行了校对整理工作，中国档案出版社罗京女士为本书的出版付出了辛勤的劳动，在此表示诚挚的感谢。

本书的出版得到了郑州航空工业管理学院信息科学系领导的支持，特别是系主任刘永教授的热情鼓励和大力帮助，这是本书得以完成和顺利出版的重要保证，在此向刘永教授表示深深的谢意！

李泽锋

2005年11月16日

于郑州航空工业管理学院

目 录

第一章 数字档案馆概论

- 第一节 数字档案馆的产生背景····· 1
- 第二节 数字档案馆的概念与作用····· 8
- 第三节 数字档案馆的功能····· 13
- 第四节 与传统档案馆、数字图书馆的关系····· 16
- 第五节 数字档案馆对档案工作的挑战····· 23

第二章 总体设计与网络平台建设

- 第一节 数字档案馆建设过程····· 34
- 第二节 数字档案馆建设的总体规划····· 39
- 第三节 网络平台建设····· 44
- 第四节 存储系统····· 48

第三章 数字档案馆综合管理系统

- 第一节 综合管理系统概述····· 61
- 第二节 电子文件收集与归档子系统····· 68
- 第三节 档案管理子系统····· 71
- 第四节 档案发布与利用子系统····· 75
- 第五节 系统管理与数据维护子系统····· 77
- 第六节 综合管理系统的开发与评价····· 80

第四章 档案网站建设

- 第一节 档案网站的功能与类型····· 89
- 第二节 档案网站建设····· 92
- 第三节 档案网站建设存在的问题与对策····· 97
- 第四节 档案网站的个性化服务····· 103
- 第五节 档案网站的评价····· 112

第五章 数字化档案信息资源建设	
第一节 数字化档案信息的来源	116
第二节 馆藏传统档案的数字化	119
第三节 档案数据库的建设	124
第四节 文字型电子文件归档的格式	128
第六章 元数据管理	
第一节 元数据概述	137
第二节 几种元数据标准介绍	142
第三节 元数据的获取与管理	151
第七章 电子文件管理	
第一节 电子文件概述	155
第二节 电子文件的管理原则	164
第三节 电子文件的鉴定与归档	169
第四节 电子文件的保管	181
第五节 电子文件的检索与利用	189
第六节 电子文件的法律效力	199
第八章 数字档案馆信息安全	
第一节 信息安全概述	206
第二节 数字档案馆网络安全	211
第三节 数字档案馆的日常安全管理	224
第九章 数字档案馆的知识挖掘	
第一节 知识管理概述	230
第二节 数字档案馆实施知识挖掘的可行性	235
第三节 数字档案馆实施知识挖掘的技术	241
第四节 利用档案信息进行知识挖掘的分析	249
主要参考文献	253

第一章 数字档案馆概论

档案并非是自人类出现就同时诞生的一种社会现象和物质实体,而是人类社会发展到一定阶段的文明产物。纵观档案的发展历程,其载体经历了由甲骨到金石、简牍、缣帛,再到纸张、缩微胶片的几次重大变革,其用途由占卜、礼器、为封建统治阶级独用到现代的为全社会所利用,每一次变革都反映了社会生产力水平的提高和社会文明的进步。随着信息化社会的到来,计算机的出现与普及,信息技术的蓬勃发展,尤其网络通信技术的广泛应用,档案载体从纸张到磁盘、光盘,出现了大量的电子文件,对档案的利用也打破了档案份数、时间和空间的限制,这一切无不在昭示着档案的又一次重大变革,又一次飞跃。为适应档案的这次变革,档案的主要保存地——档案馆也必将发展成为在网络环境下生存与发展的数字档案馆。

第一节 数字档案馆的产生背景

提起档案馆,人们首先想到的是身处政府大院的大楼和汗牛充栋的档案库。20世纪末,随着计算机技术、数字存储技术和网络通信技术的飞速发展和有机结合,尤其是网络技术在办公、生产等组织业务活动中的广泛应用,出现了新型的以磁记录为载体、以编码形式存在的电子文件。由于电子文件的复制与可在网络传输的特性,它打破了传统档案的份数和时空的限制。为了更好地对数字化档案信息进行保管和利用,一种新型的基于网络的档案馆——数字档案馆应运而生。

一、数字档案馆产生的历史背景

三千年前的甲骨档案被妥善收藏于窖穴中,金石档案珍藏于宗

庙。随着社会经济、文化的逐渐发展,国家政府机关的职能逐步明确及其强化,为了维护国家或特定集团的利益,集中管理国家或特定集团重要档案财富的各类档案机构不断涌现,如“天府”、“兰台”、“后湖黄册库”、“内阁大库”、“皇帝档案馆”、“封建庄园档案馆”、“城市档案馆”等。解放后档案工作迎来了黄金发展时期。各种类型的档案馆纷纷建设并投入使用。各级综合档案馆以及某些中央专业档案馆可归入公共档案馆范畴;而大量现存的专业档案馆、部门档案馆、企事业单位档案馆以及行将出现的社团组织档案馆、私人档案馆等均归入非公共档案馆范畴。

20世纪八九十年代是我国档案馆的重要发展阶段。从总体上看,20世纪80年代以前的档案馆以收藏纸质档案为主,通过卡片式或书本式目录或指南反映馆藏,这时档案馆业务工作的开展基本上以手工为主。

20世纪80年代以后,计算机开始进入档案馆,出现了机读式检索工具与一定数量的机读档案,计算机在档案馆中的应用为档案馆进入自动化阶段揭开了序幕。这时档案馆对于计算机的运用还主要以单机形式为主,通过功能简单的档案管理系统完成档案的著录与检索借阅工作,大量的工作还是以手工为主。

20世纪90年代中后期,以计算机技术、数据存储技术、网络通信技术为核心的信息技术全面渗透到档案馆领域,将档案馆推向自动化发展阶段。这时相当一部分档案馆尤其是大型综合档案馆和大型企业档案馆建起了馆内局域网,用户可以通过局域网访问档案馆的机读目录、光盘数据库和检索系统。一小部分档案馆实现了与互联网的连接,大量的档案目录数据库在该时期建立起来,为档案馆馆藏档案的网络化服务奠定了基础。

21世纪,互联网在我国迅速发展,渗透到政治、经济、文化等社会的各个方面。越来越多的档案馆连接到互联网,并建设了自己的网站,将馆藏档案的目录数据库直接上网。一些档案馆实现了极小部分馆藏档案的全文上网,从而彻底打破了传统档案利用的份数与

时空限制。用户在档案馆、在家中、在办公室,甚至在世界任何角落,只要有互联网的地方都可以通过网络在档案馆中查寻自己需要的档案信息。档案馆通过功能完善的档案管理系统基本实现了档案管理工作的档案编目、整理、检索、编研、鉴定等环节的自动化,这时的档案管理系统更多的是基于网络的多用户系统。

直到 21 世纪初,档案馆自动化的显著特点是用计算机(无论单机还是联网)进行数据处理,管理模式建立在手工管理的基础之上。档案管理软件主要管理的是经过著录标引得到的二次文献和经过扫描得到的档案材料的数字文本,是档案馆手工管理的辅助系统,它强调的是档案馆内部业务工作的自动化和检索工具编制和利用的自动化,而不是档案馆馆藏档案信息本身存储与检索的信息化。

近年来,计算机技术、网络通信技术、高密度存储技术、多媒体技术的高速发展、不断突破及其有机结合,档案馆自动化开始向高级阶段发展,向着数字化、网络化、虚拟化的深度和广度发展。

与此同时,档案界对许多新名词、新术语、新概念、新理论表现出了极大的热情和兴趣,如“数字化档案馆”、“虚拟档案馆”、“数字档案馆”等,这些内容已经成为档案学术期刊、学术会议、档案网站的热门话题。

随着电子政务的实施,组织局域网的铺设及网络办公的推广,产生了大量的电子文件,对这些电子文件的收集、归档、管理成了档案馆迫切需要解决的问题。传统的档案馆、传统的档案馆自动化在解决这类问题上遇到了很大困难。

1998 年我国开始实施政府上网工程,拉开了中国电子政务建设的序幕。电子政务一个很重要的功能是办公电子化、网络化。办公电子化、网络化与数字档案馆的建设息息相关。“十五”《全国档案信息化建设实施纲要》中明确提出:依托当地电子政务建设工程,建立为各级党政机关服务的档案目录信息中心,建设示范性数字档案馆。

电子政务是国家基础信息建设项目,在建设之初就应该将数字

档案馆建设纳入到总体规划之中,这既是中国档案事业的巨大机遇,也是电子政务成败的关键。

总之,档案馆需要将电子文件收集归档,并通过网络向社会提供利用服务,这就是数字档案馆产生的历史背景,也是时代赋予档案馆的任务。

二、数字档案馆产生的技术基础

20世纪80年代以来,光电技术迅猛发展,并开始得到广泛应用,光纤通信逐渐成为通信传输的主要手段,为网络的铺设奠定了物质基础。

局域网、Intranet、Extranet、Internet等各种网络渗透到社会各行各业,被广泛应用。网络将信息库中的信息以数字形式记录、编辑、组织,作为电子文档存储,同时提供浏览、访问、交流等服务。同时基于网络的应用技术也在不断创新,由C/S两层结构向B/S三层和多层体系结构发展,基于JAVA的Web Service架构设计的系统具有很好的跨平台性和兼容性。

多媒体技术是把文字、声音、图像等不同的媒体结合起来,使计算机能同时处理和综合控制这三种媒体的集成应用。多媒体将实现人们对业务活动过程的全方位记录,档案馆对其的收集、管理与提供利用使档案馆的馆藏结构发生了重大的变化。

数据存储技术是近年来又一个飞速发展同时又被广泛利用的信息技术。数据存储技术是对信息资源的整理、保存、管理和利用的最有效手段,是提高管理水平和效率的核心,也是组织信息化的核心。早期的关系型数据库(RDBMS)主要是对离散数据的管理,不很适合档案全文的存储与管理,馆藏档案目录数据库基本都是关系型数据库,这在一定程度上限制了档案信息的数字化。近年来,对象数据库得到了广泛应用,它可以将一份档案信息作为一个整体对象存储在数据库中,同时数据库的存储容量达到了TB级的海量存储,这为档案全文的数字化奠定了存储基础。

数字化档案信息由文字到图片到多媒体文件而越来越大,传统

的基于服务器的存储方法已无法满足海量存储和高速存取的需要,于是网络存储应用而生,如 NAS(Network Attached Storage,网络附加存储)和 SAN(Storage Area Network,存储区域网,简称存储域网)。尤其是 SAN,它是一种通过光纤通道共享存储设备的网络,使用专用于存储的协议,去除了网络校验等网络开销,使得网络传输速度大幅度提高,从而满足高速传输的需要。

档案全文上网是一种发展趋势。为更好地为用户服务,将不同类型的档案内容全部以数字化的方式妥善保存起来,并利用高效的查询手段对所保存的数字资源进行检索,用一定技术实现对数字内容的智能分析处理,最终使得这些数字内容能够得到最充分的利用,这就是目前蓬勃发展的内容管理技术。

社会与组织要求在其活动的过程中,把最恰当的知识在最恰当的时间传递给最恰当的人。这些知识从哪里来?相当一部分知识可以从档案馆保存的档案信息中挖掘发现,这就是知识管理技术。知识管理需要一定的技术条件,可以说没有数字档案馆的建设,档案馆提供知识服务、利用服务质量的难度是很大的。

档案馆需要与其他部门或组织或其他档案馆发生数据交换,XML 技术在这方面提供了很好的支持。

可以说,信息技术的发展要求数字档案馆的建设,同时又为数字档案馆的建设提供了技术支持。

三、数字档案馆产生的理论基础

从某种意义上看,档案馆的发展史上经历了三个发展阶段:纸质档案馆、自动化档案馆与数字档案馆。档案馆自动化也经历了三个阶段。第一阶段从 20 世纪 70 年代中期(我国在 20 世纪 70 年代末、80 年代初)到 20 世纪 80 年代后期,是档案馆自动化发展的初期阶段。美国、前苏联、英国、法国等国家在国家级档案馆开发了档案自动化系统,进行档案的自动编目与检索等管理工作。第二个阶段从 20 世纪 90 年代初到本世纪初,是局域网在档案馆得到应用的阶段。第三个阶段是档案馆自动化开始向高级阶段发展,向着数字化、网络

化、虚拟化的深度和广度发展的阶段,也就是数字档案馆阶段。这三个阶段并没有严格的界限,它们可以相互渗透,甚至现在在我国档案界还可以看到这三个阶段档案馆的具体例子。

这三个阶段中前一个阶段为后一个阶段奠定物质基础,进行技术储备。如第一阶段建设的机读目录数据库为第二阶段和第三阶段的网站检索提供了目录数据支持,第一阶段的自动编目、自动检索则为后阶段的自动著录、自动标引、档案信息检索系统的开发提供了技术积累和经验教训。第二个阶段档案馆的网络办公、网络环境下档案管理的开展为档案工作从局域网扩展到组织内网(如电子政务网、企业内网)再到国际互联网积累了经验。该阶段开展的馆藏传统档案的数字化工作则为数字档案馆下的档案全文上网奠定了基础。第二阶段信息技术的发展及在档案馆中的应用为数字档案馆技术走向实用做好了准备。

从上个世纪末,档案学术界陆续发表了多篇关于数字档案馆的论文。由于数字档案馆(Digital Archives)这一名词近几年才出现,早期的研究很少采用这个名词,而是用“电子档案馆”、“虚拟档案馆”等名词,但不管从其理论研究还是实证分析,实质上已经在开展数字档案馆的研究工作了。2003年深圳市档案局李国庆局长根据深圳市数字档案馆建设的经验主编了《数字档案馆概论》一书,对数字档案馆进行了较为系统的阐述。

目前档案学术界对数字档案馆概念的认识取得了一定的共识,形成了两种观点,对数字档案馆的功能与模式、建设方案、关键技术、安全性等也进行了分析,这一切为数字档案馆的实际建设奠定了理论基础。

从国际上看,1994年美国保护与存取委员会(CPA)和研究图书馆协会(ARL)共同组建了数字信息档案特别工作组,负责调研与推荐能确保数字档案长期存取的方法。1996年该工作组完成《数字档案长期存取与保护的总结报告》。2000年3月美国国家档案与文件管理局局长约翰·卡林宣布该局与美国国家科学基金会共同创建一

个电子档案馆。加拿大魁北克省档案馆正在实施一项庞大的纸质档案数字化计划,计划用5年时间将2200万页纸质档案数字化。

四、数字档案馆产生的实践背景

从国际上看,许多国家和地区充分利用互联网资源,纷纷建设了各种数字档案馆。上个世纪90年代初,随着计算机技术、网络通信技术的飞速发展,国外发达国家加紧了对数字档案馆的研究与开发,取得了很大成效,实际建设成功了一批数字档案馆。依据廖凯同志在《北京档案》2004年第8期发表的文章,网络调查发现数字档案馆最早出现在国际互联网上的是1992年的杰弗逊数字档案馆(Jefferson Digital Archive),由美国维吉尼亚大学图书馆设计。1992年网上可查到的数字档案馆只有两家,1997年增长到19家,2003年则达到72家。数字档案馆数量增加的同时,内容和形式上也日趋多样化。

2000年5月,“深圳市数字档案馆系统工程研究与开发”项目立项并开题研究。这是我国第一个数字档案馆项目,标志着我国数字档案馆研究进入了实际建设开发阶段。

2001年青岛市数字档案馆开始建设,并于当年完成了工程规划和基本功能的开发工作以及硬件的配置。2002年该馆开始着手大规模的数据采集工作。2003年8月20日青岛数字档案馆通过了国家档案局组织的科技成果鉴定,并于当日投入使用。

目前,各地已经完成或正在建设的档案信息化专项工程有福建省分布式档案基础数据库工程、长春市档案信息网络工程、中国第一历史档案馆网络工程等10余个。

不仅一些大城市建设数字档案馆,一些经济发达区县也加入到数字档案馆建设的行列,如杭州市萧山区数字档案馆(<http://www.archives.xs.zj.cn/>)、天津市泰达区数字档案馆(<http://archives.teda.gov.cn/ArchivesWeb/>)。

天津市泰达区数字档案馆包括两个方面:一方面,在开发区OA办公系统的基础上,解决开发区各立档单位形成文件的即时归档和利用,建成电子文件处理中心。它是一个存储和利用档案信息资源

的信息空间,是一个由众多立档单位组成的档案资源库群、档案信息资源处理中心、档案用户群构成的互连互通的虚拟档案馆。另一方面,对电子文件进行采集、整理、存储、检索、传递、保管、保护、利用、鉴定、统计等工作,是一种信息环境和基础设施。在数字档案馆建设中,电子文件处理中心是其核心组成部分,负责收集、整理、鉴定、保管、提供利用现行电子文件,数字化处理非数字化文件,并实现电子文件归档。

第二节 数字档案馆的概念与作用

一、数字档案馆概念分析

20世纪80年代出现了“电子档案馆”、“虚拟档案馆”的概念。电子档案馆的出现与当时电子文件、电子档案的最初出现密不可分。数字档案馆的概念出现于20世纪90年代中后期。随着技术的发展与完善,两者的区别逐渐淡化,趋同性特征越来越明显。因此,电子档案与数字档案虽然称谓不同,但却是一脉相承的,现在基本上已不再提“电子档案馆”,而被“数字档案馆”所取代。

虚拟档案馆的出现则是与网络通信技术的发展,尤其是互联网在中国的蓬勃发展分不开的。薛匡勇同志认为虚拟档案馆“是虚拟技术与网络技术的结合,简言之,它是通过网络技术和特定软件,将时空上分散的档案馆、室连结一体,使之成为一个有机系统。”池晓波同志则把虚拟档案馆看作“是建立在现代信息技术的全面应用基础上,以数字化方式存在的开放式档案馆。”

从本书前面的分析并对照上文关于“虚拟档案馆”的定义可以看出,无论从功能上还是技术上,今天的数字档案馆都已经将虚拟档案馆包含在内了。

近年来,国内对数字档案馆的研究和实践取得了一定的进展,但就数字档案馆的概念在理论上还没有一个统一的认识。目前在国内档案学术界主要有以下几种描述: