



SHUZIDANGANBUAN
JIANSHI
LILUNYUSHIJI

朱小怡 等编著

数字档案馆建设 理论与实践

华东师范大学出版社



朱小怡 等编著

数字档案馆建设 理论与实践

华东师范大学出版社

SHUZI
DANGANGUANJIAN
SHE
LILUNYUSHI
JIAN

图书在版编目(CIP)数据

数字档案馆建设理论与实践 / 朱小怡等编著. —上海:
华东师范大学出版社, 2007. 7
ISBN 978 - 7 - 5617 - 5469 - 6

I. 数… II. 朱… III. 数字技术—应用—档案馆—研究
生—教材 IV. G270. 7

中国版本图书馆 CIP 数据核字(2007)第 107548 号

数字档案馆建设理论与实践

编 著 朱小怡 等
项目编辑 张继红
文字编辑 张 莉
责任校对 郭绍玲
封面设计 黄惠敏
版式设计 蒋 克

出版发行 华东师范大学出版社
社 址 上海市中山北路 3663 号 邮编 200062
电 话 021 - 62450163 转各部 行政传真 021 - 62572105
网 址 www.ecupress.com.cn www.hdsdbook.com.cn
市 场 部 传真 021 - 62860410 021 - 62602316
邮购零售 电话 021 - 62869887 021 - 54340188

印 刷 者 常熟市文化印刷有限公司
开 本 890×1240 32 开
印 张 6.5
字 数 175 千字
版 次 2007 年 8 月第一版
印 次 2007 年 8 月第一次
印 数 3 100
书 号 ISBN 978 - 7 - 5617 - 5469 - 6/G · 3201
定 价 25.00 元

出 版 人 朱杰人

(如发现本版图书有印订质量问题, 请寄回本社市场部调换或电话 021 - 62865537 联系)

1	第一章 数字档案馆概述
1	第一节 数字档案馆的产生
1	一、数字档案馆的出现
2	二、数字档案馆的特点与定义
7	第二节 数字档案馆的发展
7	一、第一阶段——数据型(data)数字档案馆
8	二、第二阶段——信息型(information)数字档案馆
9	三、第三阶段——知识型(knowledge)档案馆
10	四、数字档案馆各发展阶段间的关系
11	第二章 数字档案馆的结构内容
11	第一节 数字档案馆的组成要素
11	一、数字化资源
30	二、自动化工作流程
31	三、网络化存取
33	四、分布式等第制管理
35	第二节 数字档案馆技术框架
35	一、K-W 框架的由来
36	二、数字图书馆技术框架(K-W)的介绍
37	三、狭义数字档案馆的技术框架
60	四、广义数字档案馆的建设
61	第三节 数字档案馆的社会组织形式
62	一、广义数字档案馆

62	二、狭义数字档案馆
64	第四节 数字档案馆关键技术
64	一、存储介质的有关问题探讨
79	二、安全问题
84	第三章 数字档案馆的研究类型
84	第一节 技术主导型数字档案馆
85	第二节 资源主导型数字档案馆
85	第三节 服务主导型数字档案馆
87	第四章 数字档案馆的其他相关问题
87	第一节 数字档案馆的社会功能
87	一、数字档案馆对传统档案馆的变革
89	二、数字档案馆对传统档案馆的继承和发展
90	第二节 数字档案馆与知识产权保护
90	一、数字档案馆的域名和网页的知识产权保护
91	二、数字档案馆文献资源数字化过程中的知识产权保护
92	三、数字档案馆数据库的知识产权保护
95	第三节 数字档案馆与电子政务
95	一、电子政务与电子政务信息
96	二、电子政务信息与数字档案馆资源建设
99	三、数字档案馆与电子政务的协调发展
100	第四节 数字档案馆与电子商务
102	一、电子商务与电子商务资源库建设
103	二、数字档案馆与电子商务的同步发展
104	第五章 国内数字档案馆案例介绍
104	第一节 青岛市档案馆数字档案馆调研部分
104	一、青岛市档案馆提出的数字档案馆概念

104	二、数字档案馆建设目标
105	三、建设进程与投入经费
105	四、建设成果
109	第二节 深圳数字档案馆调研部分
109	一、深圳档案馆提出的数字档案馆的概念及含义
110	二、数字档案馆总体目标和投入或拟投入经费
111	三、数字档案馆的分类
111	四、数字档案馆的主要功能
113	五、建设目标
114	六、建设的主要内容
115	七、分期规划及完成情况
117	第三节 绍兴市档案局(馆)信息化内容
117	一、工作内容和目标
118	二、项目概况
121	第四节 萧山区数字档案馆
121	一、完成全部馆藏档案的文件级条目录入工作
121	二、建立数字档案工作室
122	三、确定量子软件为萧山区数字档案馆的应用平台
123	第六章 国外数字档案馆案例介绍
123	第一节 概述
123	一、数字档案馆的理论研究项目
124	二、服务型数字档案馆的实际应用
125	第二节 永久保护真实的电子文件国际研究项目
125	一、项目背景
126	二、项目研究概况
130	三、项目研究特点
132	第三节 数字档案馆模型研究
132	一、CEDARS的产生背景

- 132 二、CEDARS 的研究概况
- 136 三、CEDARS 的研究特点
- 142 第四节 维多利亚电子文件策略项目
- 142 一、项目开展的背景
- 143 二、项目概况
- 144 三、项目的特点
- 147 第五节 美国的电子文件档案馆
- 148 一、功能定位
- 149 二、收集范围
- 149 三、保存方式
- 150 四、信息提供利用模式
- 151 第六节 英国的联合王国数字档案馆计划
- 151 一、功能定位
- 152 二、收集范围
- 152 三、保存方式
- 153 四、信息提供利用模式
- 154 第七节 加利福尼亚大学历史档案馆
- 154 一、功能定位
- 155 二、收集范围
- 156 三、保存方式
- 156 四、信息提供利用模式
- 156 第八节 杰斐逊专题数字档案馆
- 157 一、功能定位
- 157 二、收集范围
- 158 三、保存方式
- 158 四、信息提供利用模式
- 159 第九节 日本关西数字档案馆项目
- 159 一、功能定位
- 160 二、收集范围

160	三、保存方式
160	四、信息提供利用模式
161	第十节 美国 911 数字档案馆
161	一、功能定位
161	二、收集范围
162	三、保存方式
162	四、信息提供利用模式
163	附录 数字档案馆建设实施方案(试行稿)
194	本书注释
197	后记

第一章 数字档案馆概述

第一节 数字档案馆的产生

一、数字档案馆的出现

数字档案馆的构想和计算机技术的发展密不可分。早在 20 世纪 70 年代中期,随着计算机技术的广泛应用,国外的部分档案馆就已经开始了自动化之路。一方面,计算机代替手工完成了大量档案业务工作,档案馆传统的卡片式目录管理开始过渡到机读目录管理,对档案全文的信息管理也随之提上了日程;另一方面,随着办公自动化和各行各业的计算机化,接收、管理、利用电子文件和电子档案的重任也责无旁贷地落到了档案工作者身上,对档案馆的信息管理提出了挑战。同时,随着数字图书馆已经从概念变成现实,数字档案馆终于也揭开了神秘的面纱,成为 20 世纪 90 年代以来档案信息化建设的一个新的亮点。

1994 年 1 月, Margaret Hedstrom 在第二届国际人文大会上提交了《电子档案馆——网络环境的集成与利用》(Electronic Archives: Integrity and Access in the Network Environment)一文,率先提出了“电子档案馆(Electronic Archives)”的概念。1994 年 12 月,美国研究图书馆团体与美国国家档案与文件局合作成立了数字档案的特别工作组——Task Force,成员包括美国众多著名高校的教授、国会图书馆的管理人员以及 IBM 研究中心和贝尔通讯研究中心的计算机技术研究人员。1996 年,Task Force 出台了一份关于数字档案的最终报告和和建议,阐述了数字档案馆与数字图书馆在信息管理上的显著不同:数字档案馆必须保证信息的长期保管和存取利用(long-term storage and

access), 数字图书馆则没有这个责任。^[1] 这个报告设想了一个国家数字档案系统, 并从项目发起、资金保证、政策协调等方面对数字信息的长期保存和利用提出了建议。同年, 在北京举行的第 13 届国际档案大会上, 美国《档案与博物馆信息》杂志编辑戴维·比尔曼发表了《虚拟档案》一文, 提出了物理档案 (Physical Archive)、实态档案库 (Physical Repositories of Archives) 和相对应的虚拟档案、虚拟档案馆 (Virtual Archives) 的概念。从此, 正如数字图书馆发展初期一样, 我们看到电子档案馆 (Electronic Archives)、虚拟档案馆 (Virtual Archives)、没有围墙的档案馆 (Archives without Wall)、全球档案馆 (Global Archives)、超级档案馆 (Hyper Archives)、数字档案馆 (Digital Archives) 等新名称和相关的研究成果不时见诸报端。虽然在认识上依然有所差异, 但是, 随着数字档案馆概念被广泛接受, 档案界关于数字档案馆的提法已经渐趋统一。

二、数字档案馆的特点与定义

“数字档案馆”这一概念自出现以来, 始终就没有一个统一、明确的定义。这是因为, 对于数字档案馆这样一个正处于不断发展变化的新生事物, 追求一个科学完整而又公认的定义还为时太早。就现阶段而言, 对于不同的人来说, 数字档案馆意味着不同的事。凡是参与了数字化信息资源建设的群体和项目都有各自对数字档案馆的定义。档案界、计算机界、国际标准组织机构和以解决技术问题为目的的数字档案馆项目等, 都试图从不同的角度阐述这一问题。

由于数字档案馆的提法源自国外, 有必要先了解一下国外有关数字档案馆概念的论述。在国外, 数字档案馆 (Digital Archives) 一般指的是互联网上的某些站点, 具体可分为: 在网络环境下对数字档案进行存储、展览和提供利用的虚拟档案馆与传统档案馆建立的数字资源保存与利用系统两种类型; 后者主要功能是系统、全面地保存和提供政府文件系统中产生的电子文件和传统档案的数字副本。由于 Archives 一词也可用于档案资源, 因此, Digital Archives 一词也同时用来表示

网络中的“数字档案资源”。

在国外,数字档案馆建设项目大多是由国家牵头或者多个机构联合开展的合作研究。以澳大利亚公共文件署开发的用于保存和利用澳大利亚政府数字文件的数字档案馆项目为例,该馆被描述为一个可对数字对象及相关元数据进行管理和检索的信息系统,内容涉及技术构造、标准和政策三个层面。一个能对数字对象及与之相关的元数据进行管理和检索的数字仓库是其重要的组成部分。建成后的数字档案馆系统具有文件迁移、文件储存、数据管理、行政管理和业务管理五大功能。另一个典型的例子是美国国家档案与文件署(NARA)的电子档案馆建设项目,其内容包括:

1. 国家档案与文件署的网站建设;

2. 建立国家档案与文件署档案信息导航系统(英文简称 NAIL)。档案信息导航系统是拥有全国各种档案馆藏信息的联网可检索数据库,将国内所有已数字化的档案资源,按地区、来源和利用对象分别纳入统一的档案管理网络中,实现全国数字化档案资源的网上集成、查询、检索和利用;

3. 以档案信息导航系统为核心,发展一个更大的电子查询系统(Electronic Access Project),通过此系统建立一个全国馆藏档案的在线目录。其中,电子查询系统是电子档案馆最核心的部分,建立该系统主要有三个目的:第一,对 200000 条国家档案与文件署全国馆藏档案进行数字化;第二,提升国家档案与文件署的公共查询服务能力;第三,建立一个国家档案与文件署全国馆藏档案的在线目录,包括国家和地区档案馆、总统图书馆的馆藏。^[2]

在国内,关于数字档案馆的概念描述主要有以下几种:

1. 数字档案馆是指以二进制编码的数字方式存储、处理档案信息内容,应用计算机、通讯和多媒体技术,提供电子网络检索和服务的档案信息系统。它的含义有以下几点:

- (1) 接收办理完毕的电子文件目录,并对立档单位的电子文件流程实施在线监督和控制,以及时捕获电子文件,防止重要文件的流失;

(2) 将现有馆藏档案数字化,实现数字化档案资源在网上的发布和传送;

(3) 支持对馆藏各种档案实体的自动化管理;

(4) 组织对数据的有效访问和查询,使利用者可以通过网络对数字化档案信息资源进行查阅(包括目录、索引和全文);

(5) 以网络连接不同档案馆的数字化馆藏,能够提供分散于不同地区的档案信息资源,实现各个馆藏资源的无缝链接。^[3]

2. 数字档案馆是一个电子化信息的仓储,能够存储大量各种形式的信息,用户可以通过网络方便地访问它,以获取这些信息,并且其信息存储和用户访问不受地域限制。它能把包括多媒体信息在内的各种信息的数字化、存储管理、查询和发布集成在一起,使这些信息得以在网络上传播,从而最大限度地利用这些信息。^[4]

3. 数字档案馆是一个数字化的信息系统,它将分散于不同载体、不同地理位置的信息资源以数字化方式存储,以网络化方式相互连接,以便能方便使用,实现资源共享。^[5]

4. 数字档案馆是“利用电子网络远程获取档案文件信息的一种方式”。该含义重点突出了数字档案馆是在计算机通讯网络上对分布于各地的各种档案信息进行动态搜寻的实质。数字档案馆不是一种物理存在的实体,而是利用网络技术,将分布于不同档案机构的数字化档案信息资源,以网络化方式加以互相连接,便于及时利用,实现档案信息资源共享,其实质是形成一个有序的信息空间和资源共享的信息环境。^[6]

5. 数字档案馆是采用现代高新技术所支持的数字档案信息系统,是档案信息组织模式,代表的是一种信息环境和基础设施构建,是超大规模的、便于使用的、没有时空限制的知识信息中心。^[7]

6. 数字档案馆是建立在现代信息技术普遍应用的基础上,利用数字化手段,以综合档案信息资源为处理核心,对数字档案信息资源进行管理,通过高速宽带通信网络设施相连接和提供利用,实现资源共享的超大规模、分布式数字信息系统。^[8](深圳市数字档案馆)

7. 数字档案馆是政府信息资源的数据中心和数据交换站,它是建立在现代信息技术普遍应用的基础上,利用数字手段,以高度有序的综合档案信息资源为处理核心,以高速宽带通信网络为技术设施的超大规模、分布式数字信息系统,属于国家数字化信息基础工程。^[9]

8. 数字档案馆是运用数字设备和技术,收集、管理、利用档案信息的数字化、信息化、网络化的系统。它包括两方面的含义:一方面,它具备传统档案馆收集、管理、利用档案等基本功能;另一方面,与传统档案馆又有着明显的区别,它收集、管理和提供利用的对象不再是传统意义上的档案,而是数字化档案信息。^[10](青岛市数字档案馆)

9. 数字档案馆是传统档案馆功能的扩大,它以统一的标准和规范为基础,将有价值的馆藏信息资源数字化,并通过各种途径收集、捕获有价值的电子文件信息,在加工处理后以数字化形式进行存贮,并以智能检索技术为手段,提供统一友好的检索界面,利用先进的信息处理技术和互联的计算机网络,向用户提供多媒体数字信息服务。^[11]

10. 首先,所谓的数字档案馆不是一个档案馆,而是通过计算机网络连接在一起的档案馆群体。二是数字档案馆中的信息不仅仅是档案,还应包括未归档的各类电子文件和图书、资料,甚至是采集于实物的信息,可以说是上述综合性的数字信息的完整集合。三是数字档案馆不是封闭的档案馆信息网络,而是包含在办公自动化系统、计算机辅助设计和管理系统、公共信息数据管理系统等更为广阔的大系统中的一部分。四是数字档案馆不是仅仅为档案管理服务,而是面向全社会提供服务。^[12]

以上概念,代表着不同时期不同群体对数字档案馆的理解。尽管数字档案馆的概念是发展的,但我们依然可以从上面的概念中总结出数字档案馆的几个主要特征:

1. 数字档案馆以现代通信、计算机网络、多媒体信息技术设备为基础设施,它是一个档案馆集合的概念,是由一个地域或一个国家的众多档案馆或信息资源管理机构组成的联合体。

2. 数字档案馆的核心是数字化的档案信息资源。资源共享是数

字档案馆的首要法则。各个时期、各种类别的档案信息库及其他综合信息库是数字档案馆的基础资源。这些信息库可以由多个部门联合建立,但必须具有严格的标准和规范。

3. 数字档案馆是国家基础信息资源的出入口。这个出入口,既可以吸纳各种各样的信息,又可以按用户身份提供相应的服务,包括支持政府决策、提供司法凭证、开展大众教育等。数字档案馆的信息资源具有广泛性、综合性,而不仅仅局限于某个部门的馆藏,从而使其具有为社会提供全方位服务的能力。

4. 数字档案馆是传统档案馆功能的拓展和创新。数字时代,传统档案馆的服务迁移到互联网是数字档案馆的起码要求。互联网不受时空限制,可全天开放,任何用户只要拥有一台网络终端,便可随时随地浏览网上全球各类档案信息。互联网功能强大、灵活实用的检索工具,使检索成为一项轻松而愉快的工作。传统档案馆难以提供的零距离、全天候、全方位等服务因数字档案馆的诞生而实现,档案工作者与用户之间的信息交流方式也已经开始由单向性向交互性转变。

5. 从社会组织形式分析,数字档案馆可分为广义数字档案馆和狭义数字档案馆两种类型。广义数字档案馆指一个基于网络环境、采用面向对象分布式网状结构模式的数字档案馆群体。计算机通讯线路是这个网络结构的经纬,储存着数字化资源或者起着管理作用的服务器是这个网络的节点,而客户机则是这个网络的终点。广义数字档案馆可以被分解为一个个相对独立的狭义数字档案馆实体。每个狭义数字档案馆都拥有自己的数字档案资源和专业人员,并在以现代化的技术履行着传统档案馆功能的同时,为广义数字档案馆提供数字信息资源和技术支持。可以说,狭义数字档案馆是广义数字档案馆的基础,没有规范标准的狭义数字档案馆,广义数字档案馆就将成为空中楼阁。

这里有必要纠正一个错误的想法,即只要每个传统档案馆信息都上网,今天的互联网就是一个广义数字档案馆。事实上,无论广义数字档案馆如何发展,它也不可能是指整个互联网。这是因为互联网本身不是一种机构,虽然其信息管理和组织在网络技术上与数字档案馆具

有相同的协议和规范,但是缺乏档案信息应当具有的标准和证据性,甚至没有起码的分类和编目。因此,如果缺乏档案界业内的统一协调,即使每个传统档案馆都建成了狭义数字档案馆,广义数字档案馆的建设依然需要长久的等待。

综上所述,数字档案馆不只是一个技术意义上的信息管理系统,它以数字化资源、自动化工作流程、网络化存取、分布等第式管理为标志,是一个在各级档案行政管理机构的统一指导、协调下,分层履行各自职能的、高度自动化的新型档案信息管理机构群体,是对传统档案馆职能的继承与创新。

因此,我们将数字档案馆定义为:在分布式计算机网络环境中,利用数字技术、多媒体技术和计算机网络技术创建、获取、存储、处理、发布数字档案信息的档案馆或其他数字档案信息管理机构群体,是国家和地方各级政府信息化建设的有机组成部分。

第二节 数字档案馆的发展

数字档案馆的发展与现代社会知识经济的快速发展密不可分。计算机网络技术的发展彻底改变了人类交流知识和信息的方式,同时也对传统档案馆的信息管理模式提出了挑战。为了实现档案馆作为现代社会不可或缺的信息机构的知识服务功能,根据信息管理的结构,我们将数字档案馆的发展划分为三个阶段^[13]。

一、第一阶段——数据型(data)数字档案馆

这一阶段的数字档案馆以建立计算机辅助档案管理系统、接收办公自动化系统中的电子文件和做好业务系统数据库文件备份为标志。因档案馆所接收的数字资源来源不同,格式结构各异,给检索利用带来许多问题,研究人员提出了两种不同的管理策略:

(一) 集中式保管模式(加拿大哥伦比亚大学方案)

这种方案的管理与纸质文件的管理十分相似,强调将各类具有长

远保存价值的电子文件全部按规定移交档案馆集中保存。即在电子文件尚未失去现行效用时,由文件的生成机构负责保护文件的可靠性和真实性,在其失去现行使用价值后,由档案馆负责整理、编目和长期保管。这种模式强调的是文件生成机构和档案馆的各司其职。美国、加拿大、英国、北欧等国原则上赞同这一方案,并已经正式接收了部分机构所移交的电子文件。

(二) 分布式管理模式(美国匹兹堡大学方案)

这一方案主张电子文件应当始终由文件形成机构自行保管。换句话说,即便是进入档案阶段的电子文件也不必向档案馆移交,但可通过网络向档案馆进行虚拟归档。一般情况下,档案馆主要负责对非现行电子文件保管的指导工作,对这些电子文件进行登记并加以控制。这种模式强调的是将电子文件保存在产生它的软硬件环境中,以确保其可被识别、控制和存取。

这一阶段的数字档案馆建设主要致力于馆藏档案目录的数字化,并在此基础上尝试通过网络建立数字档案目录中心,向利用者介绍档案资源的分布情况,这是数字档案馆建设的起步和初级阶段。

二、第二阶段——信息型(information)数字档案馆

第二阶段数字档案馆的建设重点为:建立数字资源标准化体系,对办公自动化系统中电子文件和业务系统中数据库文件进行规范化著录(不仅仅是对数据本身,而且包括对数据产生的环境、数据的含义以及结构和来源等信息),对馆藏传统档案全文进行数字化加工,并通过建立索引标识、推进异构数据一门式检索,使利用者能够借助 Internet 或专网,对分散在不同档案馆的数字信息进行远程访问和存取。访问和存取的对象既可以是文本、图形,也可以是声音和图像等多种媒体的档案信息。档案工作者对各类档案数据的创建、获取、存储、管理、传输、发布做了大量的研究,提出了网络环境下电子文件与档案一体化管理的理念,并取得了一系列实质性的建设成果。比较具有代表性的有分布式数字档案原型(CEDARS)和开放档案信息系统参考模型(OAIS)。

该模型正日益被接受为数字信息长期保存系统的基本构架,并已作为ISO标准草案。模型详细阐述了数据提交、文档存贮、数据管理、存取及传播等内容,探讨了数据迁移、软件对保存信息的作用,以及档案馆之间的数据交换等问题。该模式可广泛应用于信息管理机构,如图书馆、档案馆等。

此外,美国的“NARA(美国国家档案馆网站)、ARIAS(宾夕法尼亚档案系统)、加里弗尼亚在线档案,日本的石川新情报书府,加拿大的数字收藏(Canada's Digital Collection)等,都可被看作是信息型数字档案馆在网络环境中实现并拓展传统档案馆的信息存储与远程服务功能的成功案例。

三、第三阶段——知识型(knowledge)档案馆

这一阶段是数字档案馆发展的最高级阶段和发展方向。知识型数字档案馆能够自动捕捉网络环境下的虚拟与实体档案信息,并按照一定的规则系统化和自组织化,实现管理者与用户之间的真正互动,使档案馆变为社会的大脑——具有丰富的记忆和融会贯通的应用记忆能力,为利用者提供经过加工处理的、可靠、完整、系统的档案信息。

这一目标的提出甚至比“数字档案馆”概念的出现更早。早在1993年,日本的信息处理开发中心(Japan Information Processing Development Center)就在《日本计算机季刊》第92期上明确提出了“知识档案馆项目”(Knowledge Archives Project)及建设知识档案馆应当具备的技术条件^[14]:

1. 信息(自动)采集(automation of the acquisition and collection);
2. 信息自组织(self-organization of knowledge),使采集的大量信息能够自动储存;
3. 知识(自动)创造功能(creation of new knowledge)。能够在大量检索的基础上根据需要提供可用信息以及相关建议;
4. 信息编译与传输功能(translation and transmission of knowledge),促进信息的交流与共享。