

数字档案馆研究

潘连根 著

中国档案出版社

数字档案馆研究

潘连根 著

中国档案出版社

责任编辑 /郭年

图书在版编目(CIP)数据

数字档案馆研究/潘连根著. —北京:中国档案出版社,2005.12

ISBN 7-80166-572-4

I. 数… II. 潘… 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 145459 号

出版/中国档案出版社(北京市宣武区永安路 106 号)
发行/中国档案出版社发行部(010-83171091)
印刷/北京振兴源印务有限公司
规格/850×1168 1/32 印张/10 字数/300 千字
版次/2005 年 12 月第 1 版 2006 年 1 月第 1 次印刷
定价/30.00 元

序

数字档案馆是随着我国信息化建设事业的深入发展应运而生的。我认为,它是破除已经在我国信息化建设过程中出现的无以数计的信息孤岛,整合并且集成管理整个机关、企业或者整个社会产生的并且储存在结构各异技术平台上的多种信息资源的重要基础和有力武器。然而,至今还很少有人能够清醒地认识到这一点。大多数人还仅仅就档案论档案,未能站到更高的高度认识数字档案馆。这说明,我国数字档案馆建设的理论尚处于摸索探讨之中,数字档案馆建设的实践也只是处于起步阶段。尽管我国有关探讨数字档案馆理论和实践的文章散见于各种档案学术刊物,但缺少系统的论述和总结。我国目前已出版的数字档案馆专著只有一部,即李国庆同志主编的《数字档案馆概论》(中国档案出版社2003年出版)。该书作为“深圳市数字档案馆系统工程的开发与研究”项目的理论成果,是我国第一家数字档案馆——深圳数字档案馆建设实践经验的初步总结。该书虽然对数字档案馆的理论作了一定程度的系统概括,但囿于当时的研究现状,理论深度稍显不足,更主要的特点是强调了自身建设中所采用的技术措施,因此该书的技术色彩浓厚,有关技术的论述占据了很大篇幅。

而潘连根同志的《数字档案馆研究》一书,在借鉴已有研究成果的基础上,对数字档案馆的理论进行了比较系统的研究。主要内容包括数字档案馆的基本认识问题、数字档案馆建设的总体论述、数字档案馆网络系统的构建问题、数字档案馆的技术问题、数字档案馆的标准化建设问题、数字档案馆信息资源的建设问题、数

字档案馆与传统档案馆和数字图书馆的比较问题等,已初步构筑起了数字档案馆的理论体系框架。可以说,本书主要侧重于对数字档案馆自身的理论研究,是对现有理论成果的系统集成,当然其中也融入了作者本人大量的有关数字档案馆的理论思考和建设建议。在某种程度上,本书与李国庆同志主编的《数字档案馆概论》可以起到互补作用,即本书偏重于理论特色,而《数字档案馆概论》一书偏重于技术特色。当然,由于时间仓促和我国迄今为止仍然缺乏真正可资总结的实践经验,本书对某些问题的研究还有待深入;对某些问题的认识(比如对元数据在数字档案馆建设中的重要作用等)也有待进一步深化。尽管如此,相信本书的出版对我国如火如荼的数字档案馆建设实践活动有一定借鉴作用,也对我国档案学理论的建设,特别是对“数字档案馆学”分支学科的构建有所贡献。

何嘉荪

2005年10月28日

前言

现代信息技术的飞速发展和网络的日益普及,对人类社会的方方面面产生了前所未有的深刻影响,也对传统的档案和档案工作产生了巨大的影响。作为对现代信息技术革命的回应,我国档案界适时提出了档案信息化建设的战略任务,而数字档案馆又是档案信息化建设的一项重要内容。

受我国数字图书馆建设的影响和启发,我国数字档案馆概念的提出与数字档案馆建设的实践基本上是同步的,从而使数字档案馆的理论研究和实践建设处于“边干边学、边学边干”的状态,在一定程度上可以说,我国数字档案馆建设实践的理论准备是不足的。

通过近几年来数字档案馆理论研究和建设实践的探讨和摸索,现在已经到了可以初步总结数字档案馆建设实践中所取得的经验教训和将已有理论研究成果进行初步系统集成的时候了。如此既有利于对今后的数字档案馆建设实践起到指导作用,也有利于在理清现有研究成果的基础上进一步深化发展数字档案馆的理论研究。

应该说,本书就是笔者在充分借鉴学界同仁大量研究成果的基础上,为此所做的努力和尝试。

目录

第一章 认识数字档案馆	(1)
第一节 数字档案馆产生的现实背景、思想理论 基础及建设意义	(2)
第二节 数字档案馆与相关概念的关系	(40)
第三节 数字档案馆的定义及特征	(49)
第四节 数字档案馆的功能及类型	(58)
第五节 数字档案馆认识中存在的若干误区	(65)
第二章 数字档案馆建设概论	(71)
第一节 数字档案馆建设的指导思想、原则和内容	(71)
第二节 数字档案馆建设的社会人文环境条件	(76)
第三节 数字档案馆的组织形式、系统模式和 运作方式	(83)
第四节 数字档案馆的评价体系	(92)
第五节 数字档案馆建设的现状及其存在的问题	(109)

第三章	数字档案馆网络系统的构建	(122)
第一节	数字档案馆网络系统的设计原则	(122)
第二节	数字档案馆网络系统的构建模式	(125)
第三节	数字档案馆智能化系统的建设	(133)
第四章	数字档案馆的技术	(141)
第一节	数字档案馆的核心技术	(141)
第二节	数字档案馆的安全技术	(150)
第三节	数字档案馆的元数据编制技术	(163)
第五章	数字档案馆的标准化建设	(177)
第一节	数字档案馆标准化建设的意义	(177)
第二节	数字档案馆标准化建设的原则	(180)
第三节	数字档案馆标准体系的构成	(182)
第四节	数字档案馆标准化建设中存在的 问题及对策	(193)
第六章	数字档案馆信息资源的建设	(198)
第一节	数字档案馆信息资源建设的含义与作用	(198)
第二节	数字档案馆信息资源建设的原则	(201)
第三节	数字档案馆信息资源的采集	(205)
第四节	数字档案馆信息资源的组织	(213)
第五节	数字档案馆信息资源的开发利用	(215)

第七章	数字档案馆与传统档案馆、数字图书馆的比较	(220)
第一节	数字档案馆与传统档案馆的比较	(220)
第二节	数字档案馆与数字图书馆的比较	(229)
第八章	数字档案馆若干相关问题的研究	(240)
第一节	数字档案馆的著作权保护	(240)
第二节	数字档案馆的隐私权保护	(253)
第三节	数字档案馆的知识管理	(263)
第四节	数字档案馆的研究	(275)

第一章 认识数字档案馆

众所周知,档案馆是人类社会发展到一定阶段的产物,并随着社会的发展而发展。社会的发展和技术的进步必然推动档案馆的发展和进步,社会发展水平与技术水平决定着档案馆的发展水平,这已为中外档案工作发展的历史和现状所证实。目前,数字档案馆顺应社会发展和技术进步的潮流已应运而生。但数字档案馆作为一项系统工程,在建设过程中,我们面临着许多问题,需要研究解决。其中,首要的问题便是如何认识数字档案馆这一新生事物,因为对这一问题的不同理解,将直接影响到数字档案馆的具体设计和实施,甚至会使数字档案馆的建设走上弯路。这方面,数字图书馆在建设过程中就曾经有过这样的教训,应该成为我们的前车之鉴。“美国数字图书馆建设早期,主要依靠计算机通信领域的专家和技术人员具体实施,科学研究和工程建设几乎完全脱节,所以其研究成果有些不切实际;法国国家图书馆在数字化建设中,由于混淆了传统图书馆与数字图书馆的区别,设计、实施全由图书馆自己承担,结果出现问题,只有推倒重来,造成巨大浪费。”^①可见,对数字档案馆的正确认识和理解,不仅直接关系到数字档案馆的实践,即建设的是一个什么样的数字档案馆,而且自身又是一个重要的理论问题。本章主要通过分析数字档案馆产生的现实背景、

^① 田捷编著:《数字图书馆技术与应用》,科学出版社2002年版,徐文伯序。

思想理论基础及建设意义,阐述数字档案馆与相关概念的关系以及数字档案馆的定义、特征、功能、类型,并进一步剖析数字档案馆认识中存在的误区,以期读者对数字档案馆大致有一个整体性的正确认识。

第一节 数字档案馆产生的现实背景、 思想理论基础及建设意义

数字档案馆是20世纪90年代末现代信息技术和社会发展的直接产物,它的产生既有其现实背景,也有其深厚的思想理论基础,数字档案馆的建设具有十分重要的意义。

一、数字档案馆产生的现实背景

从宏观和整体角度看,档案馆事业的发展和档案馆形态的演变,是在档案馆系统内部驱动力和外部推动力的双重作用下得以发生和实现的。因此,关于数字档案馆产生的背景,我们可以具体从数字档案馆产生的内在驱动力(内因)和外部推动力(外因)两个角度来进行分析。

(一)数字档案馆产生的外因

数字档案馆产生的外部推动力量主要源于档案馆所处信息环境的改变。

回顾20世纪百年历程,冲击和影响人类社会生活各个方面幅度最大、速度最快、震撼最强的,莫过于延续至今仍在飞速发展的信息技术革命浪潮。尤其是20世纪90年代以来,人类开始步入信息化、网络化、数字化时代,整个社会包括各种组织、个人都在经历着前所未有的全方位的变革,档案馆也毫不例外地经受着变革的洗礼。

1 信息基础设施(NII,俗称信息高速公路)的建设,是档案馆所处信息环境发生变化的重要原因,也是数字档案馆得以实现的

重要推动因素。

1993年9月,美国克林顿政府宣布实施为期20年、耗资4000亿美元的《美国国家基础设施行动计划》(即NII计划),这是美国在全面步入信息化社会之际实施的一项长远的、具有划时代意义的重大战略决策,该计划的中心内容就是要在21世纪,使各种媒体的信息集成在强大的计算机上,并能联网快速传送展示,并要求在2015年之前,使美国所有的家庭、图书馆、档案馆、办公室等都能通上光缆,从而使美国的信息化水平有一个大的飞跃,并以此推动美国综合国力的进一步提高,继续确保其第一大国的地位。

美国的NII计划引起了世界各国的巨大反响。许多发达国家和地区如英国、加拿大、欧共体、日本等不甘落后,纷纷制定了本国的信息高速公路计划,决心与美国一争高下。如日本就计划在2010年(比美国提早5年)完成光缆铺设,并在2000年以前首先在全国国立、公立的大学、医院、图书馆、档案馆铺设光缆,同时还制定了一系列的政策来鼓励民间企业参与。就是许多发展中的国家和地区,也对此做出积极反应,纷纷出台各自相应的计划。如我国在1993年年底,国务院信息化工作领导小组制定了《国家信息化“九五”规划和2010年远景目标(纲要)》,开始启动“三金”工程(即“金桥”工程、“金卡”工程、“金关”工程)，“三金”工程旨在推动我国信息化基础设施建设走向深入,立足于重点行业和部门传输数据和信息,提高工作绩效。“三金”工程的提出和实施,使我国出现了一股强大的信息化浪潮,对政府、企业和家庭产生了强大的影响,国家各有关部委以此为契机,先后提出和实施了一系列“金”字号系统信息工程,如“金税”工程、“金农”工程、“金企”工程、“金盾”工程、“金信”工程、“金水”工程、“金审”工程、“金质”工程等。

由于信息高速公路是以计算机技术、网络通信技术、多媒体技术等先进的信息技术为基础,以光导纤维、数字卫星系统等为主要信息传输载体,以高速传递和处理信息、最大限度地实现全社会信

息资源共享和高度社会经济信息化为目的,运用大容量、高速交互式的信息网络把政府机构、科研单位、公司企业、图书馆、档案馆、学校、家庭等的信息终端连接起来,从而奠定面向未来的社会信息基础设施。因此,信息高速公路的建设,使整个社会信息流通的基础结构发生了根本性的变化,从而也使档案馆所处的信息环境有了根本改变。同时,信息高速公路的建设和实施,一方面要最大限度地应用大量已有先进的信息技术,另一方面又将进一步推动信息技术的继续发展,而这些先进的信息技术恰恰为数字档案馆的建设提供了技术基础。

2. Internet 的迅速推广与普及构成了现代信息环境的第二个重要变化。它为数字档案馆的建设提供了现实的网络环境和丰富的网络信息资源。

Internet 起源于 20 世纪 60 年代的美国,它是在 1969 年美国高级研究计划署(ARPA)投资建立的 ARPANET 的基础上逐步发展起来的。ARPANET 最初使用的网络协议是“网络控制程序”(NCP),但在 1980 年被目前广泛应用的由美国国防通信局(DCA)和国防部高级研究计划署(DARPA)研制成功的“传输控制协议/网络互联协议”(TCP/IP)所取代。1985 年美国科学家基金会(NSF)通过对 ARPANET 进行重建,建立了 NSFNET,将美国的大学和科研机构的计算机网络连接起来。随着计算机、远程通信技术的发展和人们对信息交流与共享需求的增长,越来越多的网络开始连接到 NSFNET 上,人们将这个以 NSFNET 为主干的并连接了大量具有不同软硬件的计算机网络称为 Internet(在我国称之为“因特网”、“国际互联网”、“网际网”等)。可见,Internet 是网络的网络,是以广域网把无数局域网利用 TCP/IP 协议互相连接起来的网络。Internet 以其覆盖的全球性、灵活的人网方式、快速的信息传递、丰富的信息资源、低廉的费用、种类繁多的信息服务方式和广阔的发展前景等优点受到广大用户的青睐,已成为世

界上覆盖面最广、规模最大、信息资源最丰富的计算机网络。

我国在 1990 年 10 月正式向 Internet 国际互联网络信息中心 (INTERNIC) 注册了我国的最高域名“.CN", 并从 1994 年起实现了与 Internet 的 TCP/IP 连接, 开通了 Internet 的全功能服务。特别是几个全国范围内的计算机网络的相继建立, 使 Internet 在我国得到了较快的发展。目前, 我国已形成国家级六大骨干网络: 中国公用计算机互联网 (CHINANET)、中国教育和科研计算机网 (CERNET)、中国科技网 (CSTNET)、中国金桥信息网 (CHINANAGBN)、中国联通互联网 (UNINET)、中国网通 (CNCNE)。据中国信息产业部副部长奚国华在第四届中国网络媒体论坛上介绍, “中国互联网发展十年来, 其用户规模、业务应用、运营管理和技术水平等方面都发生了巨大变化。目前, 中国互联网上网人数超过 9000 万户, 居世界第二位。其中宽带接入用户达到 2000 万户, 互联网站 60 万个, 上网计算机 3600 万台, CN 域名注册总量 40.5 万个, 互联网国际出口带宽超过 60G。同时, 具有自主知识产权的网络设备、网络终端产品以及相关软件的研发和国产化水平不断提高, 基于互联网的电子商务、电子政务、远程教育、电子娱乐等新业务层出不穷, 社会信息化建设进程不断加快。”^① 特别是在 2004 年 12 月 25 日, 中国第一个下一代互联网暨中国下一代互联网示范工程核心网 (CERNET2) 正式开通。第二代中国教育和科研计算机网是“中国下一代互联网示范工程”中最大的核心网和唯一学术网, 它以每秒 10G 的传输速率 (每秒 100 亿比特, 相当于每秒传送 15 个 VCD 光盘存储的信息) 连接全国 20 个主要城市的核心节点, 为全国几百所高校和科研单位提供高速 IPv6 (互联网协议第 6 版) 网络接入服务, 高速连接国内外下一代互联网, 并且

^① 《中国信息化建设取得长足发展》,《文汇报》2004 年 11 月 16 日第 12 版。

在 CERNET2 的核心节点配置的 IPv6 核心路由器和接入路由器, 50% 以上采用的是国产设备, 标志着我国在下一代互联网关键技术上获得了突破性进展, 从而使我国在下一代互联网的发展上已经赶上了世界的步伐, 取得了先机。中国教育和科研计算机网专家委员会主任、清华大学教授吴建平认为, 下一代互联网具有以下特点: (1) 更大。目前的网络协议为 IPv4, 可提供的 IP 地址大约为 40 多亿个, 而且由美国掌握绝对控制权, 2005 年左右将全部分配完成, 全球将面临严重的 IP 地址枯竭的危机。下一代互联网将逐渐放弃 IPv4, 启用 IPv6 地址协议, 从 2 的 32 次方增加到 2 的 128 次方, 从而使 IP 地址资源极端丰富, 有人比喻为世界上每一粒沙子都会有一个 IP 地址, 可以充分满足数字化生活的需要。(2) 更快。下一代互联网将比现在的网络传输速度提高 1000 倍以上, 它的基础带宽可能会是 40G(每秒 400 亿比特, 相当于传送 10 个 DVD 影片)以上。(3) 更安全。目前的计算机网络因为种种原因, 在体系设计上有一些不够完善的地方, 下一代互联网将在建设之初就从体系设计上充分考虑安全问题, 使网络安全的可控性、可管理性大大增强。^① 基于以上特点, 下一代互联网将使人们进一步享受到网络带来的好处, 网络计算、高清晰度电视、强交互点到点视频语音综合通信、智能交通、环境地震监测、远程医疗、远程教育等都将得以真正实现, 真正的数字化生活即将来临。

网络通信系统是数字档案馆的重要基础设施, 丰富的网络信息资源也是数字档案馆“虚拟”馆藏的重要来源, 而使用户能够通过网络联机存取档案信息资源是数字档案馆建设的重要目标之一。Internet 的迅速推广和普及, 必然推动数字档案馆在档案信息资源建设和档案信息服务的广度和深度方面有一个质的提高。因此, “具有高水平的网络技术、丰富的网络化信息资源、便利而强

^① 《下一代互联网开通》,《文汇报》2004 年 12 月 26 日第 2 版。

大的网络服务功能 Internet 的发展为数字档案馆的发展提供了最理想的网络环境。”^①

3 电子文件/档案的激增构成了数字档案馆信息环境的第三个方面的变化。电子文件/档案是数字档案馆的重要物质基础,是数字档案馆得以产生的信息资源基础。

始于 20 世纪 50 年代在美国等西方国家兴起的办公自动化,在 80 年代到 90 年代得到了较快的发展。办公自动化系统是一个开放的、动态的、发展的系统,它随着信息技术的不断发展而改进。随着信息基础设施的建设和完善,Internet 的推广和普及,电子政府、电子商务、电子军事、电子货币等新的工作方式和生活方式应运而生,从而在人们的工作和生活中出现了一种全新的文件载体——电子文件。办公自动化(OA)和各种数字技术环境如计算机辅助设计(CAD)、计算机辅助制造(CAM)、计算机管理信息系统(CISM)等等的日益普及,使电子文件的数量激增,大有后来居上之势。同时,随着 1999 年“政府上网工程”的实施,档案管理部门作为政府系统的有机组成部分,也必然要参与其中,迫切需要为政府各部门及社会公众提供联机在线的电子文档服务。此外,我国于 20 世纪 80 年代开始的档案自动化工作,也已将部分传统档案进行数字化处理转换成了电子档案。这些日益增多的电子文件/档案为数字档案馆的建设提供了充足的一级文献“货源”。

总之,在全球信息化浪潮中,信息社会的基础结构正在形成。信息已经和物质、能源共同成为人类社会发展的三大资源。在日趋激烈的竞争环境中,信息意味着权利、机遇和发展,信息资源的开发利用水平直接关系到国家的社会 and 经济发展水平。但“目前占世界人口 20% 的发达国家拥有全世界信息量的 80%,而占世界人口 80% 的发展中国家却只拥有信息量的 20%。因此,对于发展

^① 金波.《为什么需要数字档案馆》,《档案学通讯》2002 年第 2 期。

中国家来说,信息资源和知识的匮乏是发展中面临的最大困惑。”^①“从国外信息化进程的成功经验看,往往是三分信息技术,七分信息设备,十二分的信息资源。”^②因此,数字化信息资源的开发利用是信息化建设的核心。我国的信息化建设尽管起步较晚,^③但发展较快,已初具规模,不过仍存在着因投入不足而造成的数据库种类少、数量少、容量小的突出问题。因此,目前我国信息化建设中最紧迫、最薄弱的环节是数字化信息资源的建设和开发利用问题。

作为国家重要公共信息资源的图书文献资源被视为国家信息基础设施建设、知识创新体系和创新能力的重要组成部分,首先开始了数字图书馆的建设实践。事实上,我国自20世纪90年代中期开始跟踪国际上数字图书馆的发展动态,着手进行技术研发。1997年国家计委批准了国家重点科研项目“中国试验型数字式图书馆计划”,项目实施期限为1997年7月至1999年12月。1998年7月,中国国家图书馆向文化部提出了“中国数字图书馆工程”的立项申请,经文化部领导研究后,决定报国务院科技领导小组,向国家计委申请立项,2000年4月,成立了以文化部牵头、由21个政府部委单位参加的“中国数字图书馆工程建设联席会议”,标

① 中国科技信息研究所、北京大学信息管理系、北京图书馆、清华大学图书馆、中国科学院计算机网络信息中心:《我国科技电子信息资源的开发和利用研究》,北京图书馆出版社1999年版,第4页。

② 同上。

③ 对于中国信息化起步于何时,主要有两种观点:(1)大多数学者认为是国家经济信息化联席会议成立的1993年(信息产业部指定“国家信息化培训教材”《中国电子政务领导干部知识读本》,中共中央党校出版社2002年版,第54页);(2)也有一些学者认为是周恩来总理亲自主持制定《十二年科学技术发展规划》的1956年(徐祖哲:《中国信息化是何时起步的》,《光明日报》2004年11月10日C1版)。