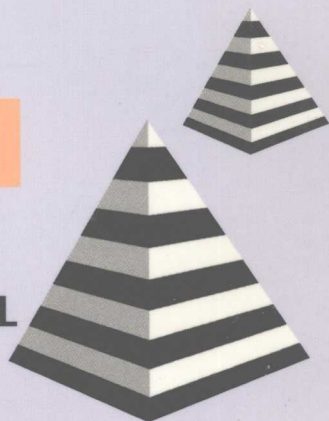


社会科学前沿

档案数据库的规范和质量控制

钱 毅 著

THE STANDARDS AND
QUALITY COTROL ON ARCHIVAL
DATABASES



中国传媒大学出版社

档案数据库的规范和质量控制

作者：钱 毅

江苏工业学院图书馆
藏书章

G270.7
QY

中国传媒大学出版社

图书在版编目 (CIP) 数据

档案数据库的规范和质量控制 / 钱毅著. —北京: 中国传媒大学出版社, 2007.5
ISBN 978-7-81085-970-7/K · 970

I. 档... II 钱... III 数据库—应用—档案管理 IV.
G270.7

中国版本图书馆 CIP 数据核字 (2007) 第 070222 号

档案数据库的规范和质量控制

作 者 钱 毅

责任编辑 王 壬

封面设计 耕者工作室

出 版 人 蔡 翔

出版发行 中国传媒大学 出版社

北京市朝阳区定福庄东街 1 号 邮编 100024

电话: 86-10-65450532 65450528 传真: 65779405

<http://www.cucp.com.cn>

经 销 新华书店总店北京发行所

印 刷 北京市德龙公防防伪印刷厂

开 本 850 × 1230mm 1/32

印 张 6.75

版 次 2007 年 5 月第 1 版 2007 年 5 月第 1 次

书 号 ISBN 978-7-81085-970-7/K · 970 定 价: 28.00 元

版权所有

翻印必究

印装错误

负责调换

自序

在档案信息化的进程中，大量的信息技术被引入到档案实践领域，必然产生许多与传统档案管理迥异的新问题。如何将信息技术与档案管理理论与实践进行有机结合，提升档案管理理论的生命力，引导适于档案管理要求的技术发展，是新时期档案学研究的重要内容。数据库作为当前档案信息资源组织的最佳形式，被绝大多数档案单位采用作为资源存储、管理的主要工具，在应用过程中出现了诸如共享困难、结构失范、质量参差等问题。

在我本人经历的档案公文网络一体化系统项目的开发过程中，曾体验过由于档案数据库结构标准缺乏而导致程序频繁改动的切肤之痛，同时在多次项目调研活动中也了解到实践部门对档案数据库的诸多诉求。深感于这个问题在理论和上的潜在价值，笔者不自量力，在国内外资料普遍缺乏的情况下坚持进行了多年的思考。在大量文献和实地的调研基础上，分析了我国档案数据库建设存在的问题及原因，将档案学的基本理论运用于数据库建设与管理技术之中，提出了档案数据库建设影响因素模型，并在此基础上对档案数据库规范和质量控制问题提出自己的见解。笔者不揣冒昧，自认为本书创新主要有三：一是从内容、业务环境、应用等多种角度强化了对档案数据库的认识，形成了全面立体的“档案数据库”观；二是从层级、数据描述、数据库建设的三维角度构建了档案数据库的标准体系，并就核心标准展开了深入的分析；三是提出了构建以标准体系为基础的档案数据库质量控制体系的概念。

莺啼初试，雏凤新声。幼稚肯定在所难免，涵望各位专家读者不吝赐教。

目 录

第1章 导论	1
1.1 研究背景	1
1.2 选题意义	3
1.3 国内外研究现状	4
1.4 论文研究设计	7
第2章 档案数据库概述	10
2.1 档案信息资源组织的三个阶段	10
2.2 档案数据库概念界定	19
2.3 档案数据库是档案信息系统建设的主要内容	24
2.4 档案数据库类型分析	29
2.5 档案数据库特点与局限	36
2.6 档案数据库现状分析	43
2.7 档案数据库建设的意义	46
第3章 影响档案数据库建设的要素分析	54
3.1 档案数据库影响因素的确定	54
3.2 档案管理理论影响因素	59
3.3 数据库基本理论支撑档案数据库的设计与实现	65
3.4 信息系统的不同属性决定了档案数据库的表现	69
3.5 档案数据库需要科学规划管理	75
3.6 组织管理	79
3.7 数据状况	83
3.8 数据库工程	86
3.9 标准规范是影响档案数据库建设的综合因素	88
第4章 档案数据库规范研究	91
4.1 档案数据库规范的意义	91

4.2	档案数据库标准建设现状分析	95
4.3	档案数据库的标准规范分析	103
4.4	构建档案数据库标准体系	151

第5章 档案数据库质量控制 175

5.1	档案数据库质量分析	175
5.2	构建以标准体系为基础的档案数据库质量控制体系	185

参考文献 203

第1章 导论

1.1 研究背景

档案作为极其重要的社会信息资源，在目前的信息环境中基本上都采用数据库的形式对其目录或内容进行管理，基于档案数据库我们可以开发各种形式的具体应用，它已经成为各类档案信息系统的核心部分，也得到了各个层面的普遍重视。

根据《中共中央办公厅、国务院办公厅关于加强信息资源开发利用工作的若干意见》（中办发[2004]34号）的精神，2005年国家档案局下发了《国家档案局中央档案馆关于加强信息资源开发利用工作的意见》（档发[2005]1号文件），其中第九条明确提出要“开发和应用统一规范的档案信息管理系统，建立馆藏档案的规范化管理和利用平台。……以各省、自治区、直辖市国家档案馆为依托，分别建立本区域范围的档案信息资源库，构建统一管理、分级共享的区域档案信息中心。逐步在全国建立起区域性、多层次、分布式、规范化的各类档案数据库群，通过各类信息网络平台实现档案信息资源共享。加强规范化建设，统一数据格式和管理系统，确保信息交换通畅和便于共享。”其中对于档案数据库的规范

性、交换性、共享性给予了高度关注。

在国家档案局和各主要省市的规划中对档案数据库也给予了足够的重视。《全国档案信息化实施纲要》中提出要加强档案目录数据库、多媒体数据库、全文数据库的建设。国家档案局十一五规划(草案)的实施项目中明确提出要“建立国家档案信息目录体系。建立中央档案馆全国革命历史档案目录中心、中国第一历史档案馆全国明清历史档案目录中心、中国第二历史档案馆全国民国历史档案目录中心为主干的国家档案信息目录体系。该目录体系以各级国家档案馆馆藏为主体对象,以各级各类档案室保存档案为补充。”以及“建立若干档案信息数据库。建立电子化的全国珍贵纸质档案全文数据库和多媒体档案数据库。开展档案数据的深加工和专题编研,建立一批直接服务经济社会发展需要的重要专题档案数据库。”同样,在各地地方档案主管机构的规划中,对档案数据库的建设也非常重视。《海南省档案信息化建设规划(2005—2010)》^①全文一万字中,“数据库”就出现了36次之多,《青岛市区市数字档案馆建设方案》中也出现了近20次。

档案数据库的战略地位得到普遍认同以及其数量、类型的发展固然可喜,但档案数据库建设过程中的许多问题依然存在。2002年,《档案学研究》第3期发表了浙江省档案数据库建设调查组关于该省综合档案馆信息数据库建设的调查报告,该报告花了很大篇幅比较详细地反映了当前档案数据库普遍存在的许多问题,包括标准规范缺失、逻辑设计混乱、质量不容乐观等。时隔几年,关于此类数据库规范与质量问题的讨论依然在进行。根据中国人民大学档案学院档案网站调研小组对我国省、自治区、直辖市、特别行政区、计划单列市、省会城市档案局(馆)网站的综合调查(2003.12~2004.5)^②,在所调查的40个网站中,能够较为稳定提

① <http://archives.hainan.gov.cn/gov/jdangan/20041222.doc> [2006-9-1]

② 《我国省、自治区、直辖市、特别行政区、计划单列市、省会城市档案局(馆)网站调查报告》,《档案聚焦》,2004年第6、7期

供档案检索和浏览查找服务的只有 17 个,其中只有两个网站能够提供部分原文供用户查看,即使提供检索服务的网站,档案信息表达都过于简单,无法为访问者提供足够详尽的介绍。调查者认为这些“与档案馆数据准备的基础工作有很大关系”,并“一致认为在档案网站的所有功能中,最需要提高、最值得提高的就是档案数据库和检索工具的建设。”

1.2 选题意义

本书尝试从档案管理和数据库技术的视角对档案数据库当前存在的主要问题进行分析,以期对档案数据库建设过程中的关键问题提供有价值的帮助。

1.2.1 本书的理论意义

首先,界定了本书的研究对象——档案数据库的基本概念。国内外的实践表明,数据库是信息资源开发的有效途径,一个国家或行业拥有多少数据库,能用数据库提供多少服务是其现代化水平的标志之一。对于档案数据库,应当用战略的眼光来审视,而不能将其仅仅视为存储档案数据的场所。档案数据库反映了档案存储、管理、服务等活动的综合需求,是档案信息及其管理信息活动的忠实记录者、管理者、保管者,通过综合考察档案管理基础、数据库建设流程、应用环境等因素,界定了档案数据库的基本概念。

其次,深化了对档案数据库的认识。作为数据库与档案信息空间交汇的产物,档案数据库综合体现了二者的特点,这些特点在不同的场合可能有不同的表现,把握双方的特点对档案类型进行区分,有助于在管理要求、标准规范等方面区别对待。本书从档案的内容、数据形式、数据结构、系统环境等角度对档案的类型进行分析,基本涵盖了档案信息化领域对档案数据库的认知。同

时也对档案数据库的特点和局限性进行了分析。

第三,综合考察了档案数据建设的影响因素。档案数据库建设是档案信息化的核心内容之一,本书综合考察了档案数据库建设的多种影响因素,分析这些因素如何影响档案数据库的表达和实现,按照范围—层面—因素的分析结构获取影响因素分析框架。通过该框架全面把握档案数据库在建设时需要正视的问题,从而也为档案数据库的建设规范和质量控制提供思路。

1.2.2 本书的实践意义

在对档案数据库进行概念界定、影响因素分析的基础上,本书力图通过构建档案数据库标准体系及主要基于该体系构建的质量控制体系,为档案数据库的建设提供具有实践意义的指导。

首先,构建档案数据库标准体系。档案数据库的建设需要大量标准规范的支撑,这些标准不是简单的数据库建设规范和档案标准的集合,而需要根据业务模式的不同,按照多个标准维度进行组织的标准体系,特别强调了标准体系的数据描述维度各标准的规范性,对于构造标准的、可开放联合、可交换的数据库具有实际指导作用。

其次,构建档案数据库质量控制体系。档案数据库质量问题是档案信息资源价值能否发挥的关键因素,良好的数据库质量是数据库稳定发展的基本条件。通过对档案数据库质量内涵、评价指标以及现状的分析,提出构建基于标准体系的质量控制体系,便于理解和实施。

1.3 国内外研究现状

1.3.1 国内研究现状

由于档案数据库建设是一项对管理基础和技术要求都较高的实践性项目,针对档案数据库的研究在实践领域和理论研究中都

不同形式的表现。

在实践领域,针对档案数据库的研究和工作内容体现在工作规划、档案信息管理系统的项目文档、档案软件商的方案设计中。这方面比较典型的是福建省《分布式档案基础数据库建设理论与实践》、《档案信息化建设实务》中介绍的浙江省档案馆目录数据库的实例,以及诸如青岛数字档案馆建设的工作报告等。实践领域对档案数据库的研究重点更多地倾向于技术层面,对数据库选型、系统配置、系统架构、数据容量、工作流程等方面的内容着墨甚多。而对如何组织数据库、引用标准、数据库管理则较少涉及。

相对而言,学术研究领域针对档案数据库的研究要薄弱得多。在期刊网(www.cnki.net)上以“档案数据库”检索1994~2005年的文献篇名,在精确检索模式下只找到28篇命中文献^①,以该词汇检索“中文摘要”字段,找到73篇文章,而以“档案数据库标准”检索“关键字”字段,只找到一篇文章。通览相关文献,了解到目前国内学术领域对档案数据库的研究表现出如下特点。(1)关注度不足。在实践部门知名度很高的数据库在档案学术领域没有获得足够的关注,特别是对档案数据库建设过程中需要的档案业务方面的理论指导严重不足。这也直接导致档案数据库的建设和研究过程中“两张皮”现象较为普遍,在实际部门关于档案数据库的工作报告中这个特点也非常突出。分工细化、知识结构差异是导致这种情况的主要原因。(2)研究内容松散,理论性普遍较差。档案数据库文献量虽然不大,但关注的范围包括安全、著录、软件应用、质量、著作权、特色馆藏等诸多方面,且文章多数是部门级档案数据库的应用心得,缺乏宏观的、理论性的指导,意见流于泛泛。

① 检索时间:2005年9月16日,检索数据库为:电子技术及信息科学辑目录。

1.3.2 国外研究现状

对国外档案数据库的考察主要通过文献研究和网络访问两种方式进行。对国外档案数据库的文献研究,笔者以“archival database”查询了PQDD (ProQuest Digital Dissertation) 博硕士学位论文数据库,无一篇命中文献。后虽放宽检索条件,命中文献也寥寥可数^①,且文献内容与笔者研究内容差异很大,于是退而关注主要发达国家档案馆实施项目和有关标准文本的研读。

通过对主要发达国家和地区档案网站的访问,了解到国外档案数据库在标准规范、内容建设、提供利用等各方面都具有很大的优势。主要参访网站包括:

□ARC: Archival Research Catalog, 档案检索目录是美国国家档案与文件署(NARA)保存在华盛顿地区,地区档案馆和总统图书馆的全国性档案的在线目录,可检索到包括 10 个总统图书馆在内的 200 多个档案馆的数字化档案。

□AAD: Access to Archival Databases, NARA 的档案数据库检索系统,包含 49 个类别(Series)下 475 个案卷(File)约 85,000,000 份历史电子文件(record)。

□NWDA: Northwest Digital Archives, 美国西北数字档案馆,提供的档案信息全部来自华盛顿、俄勒冈、爱达荷、阿拉斯加和蒙大拿五个州的档案馆,通过EAD (Encoded Archival Description) 进行档案描述编码建立的联合数据库,NWDA 提供更加强大的检索方式检索五个州的各种档案和手稿材料。

□ArchiviaNet: (<http://www.collectionscanada.ca/archivianet/index-e.html>) 加拿大图书档案馆(LAC)的档案查询

①针对PQDD的进一步检索,笔者以subject_name=%Library Science% and t_title=database and abstract=archives 检索命中1篇文献,该文是关于将缩微影像建库的研究报告;以subject_name=%Library Science% and t_title=database 进行检索命中3篇文献,其中没有关于档案数据库方面的内容。

系统,依据加拿大RAD著录标准进行各类档案描述,提供政府档案文献和其他资源的检索利用,包括文件原文、地图、艺术品、照片、录像、胶片、声音、建筑绘画、邮票以及奖牌。

□NDAD:英国公共档案馆(PRO)的联机档案数据库与检索工具,保存和提供在线的档案文件和数据接口,可查询由英联邦政府提供的自1963以来的英国近代政务信息,这些信息是基于ISAD(G)标准进行描述的。

1.4 论文研究设计

1.4.1 研究内容

论文主要包括五章。

第一章导论,对研究背景、选题意义、研究方法与创新进行了简单介绍。

第二章主要探讨了对档案数据库的基本认识。明确档案数据库是档案信息组织的高级形式,确定其概念及其基本内涵,并以此作为全文研究的逻辑起点。通过对档案数据库的类型、局限、现状、意义等方面的具体分析,明确档案数据库在档案管理活动中的定位,为全篇研究提供理论基础和实践背景。

第三章较为详尽地分析了档案数据库建设的影响因素。提出了档案数据库的影响因素模型,从信息系统、组织、社会环境三个范围域和理论、数据、规划、业务、规范五个层面搭建了档案数据库因素分析框架。通过对变量因素的优化,确定了档案基本理论、信息系统、组织规划管理、数据状况、标准规范、数据库工程等重点因素,分析了它们对档案数据库建设的影响。档案数据库影响因素分析是对档案数据库认识的进一步深化,为档案数据库标准体系和质量控制提供了分析基础。

第四章研究分析与档案数据库相关的标准规范。在明确标准规范意义的基础上,从档案业务模式、数据库系统生命周期、数

据管理层次等角度对相关标准规范进行分析,综合比较国内外的档案标准研究现状,提出了档案数据库标准体系的概念和基本框架,包括体系建设的基本原则、维度分析,以及不同业务模式下标准体系的构建要点。

第五章研究了档案数据库质量控制问题。档案数据库质量问题是目前档案数据库建设中亟待重视和解决的关键问题,明确档案数据库质量的内涵,从数据、数据表、数据库三重角度来把握。分析数据库质量评价体系,阐明各个评价指标的含义。文章认为,应当采用体系的方式来进行档案数据库的综合质量控制,分析了档案数据库与一般软件产品的不同之处,可以参照而不能照搬CMM、ISO9000族的质量管理标准,而应基于档案数据库标准体系构建质量控制体系。

1.4.2 研究方法

本书研究的对象具有较强的跨学科色彩,包括档案管理、元数据理论、数据库系统、软件工程、标准理论、质量控制等方面的内容,同时档案数据库管理也是理论性和实践性并重的管理活动。在研究方法中,本书着重采用了以下研究方法:

1. 系统法。系统分析法是本书研究中采用的主要方法,它倡导用整体、联系的思维角度来认识研究对象,有利于从广泛的领域和视角来认识档案数据库的诸多问题。档案数据库影响因素架构的确定、档案数据库标准体系和质量控制体系的构建都是依据系统论的基本思想,从整体联系的系统角度出发进行考察的结果。现代科技和管理活动的发展使得科学研究必须借鉴其他学科的理论和研究成果,学科间的交叉融合也是目前档案研究活动必须面对的现实,针对本书研究对象的跨学科色彩,在研究方法中采用系统分析的方法,将分散在各专业领域的因素进行综合思考。

2. 比较研究法。比较方法是将类似的事物进行对照,区别其异同的思维方法。通过研究类同事物之间的差异,分析其产生的

原因,加深对研究对象内在本质的认识,提出符合研究对象实际情况的措施。在数据标准的论题上,本书对国内外著录思想、采标情况、不同业务模式的适用标准、乃至不同的著录标准之间都多次采用比较研究的方式进行展开。

3. 理论联系实际。在本书论述中,运用档案管理、数据库系统、软件工程等基本理论来分析论证研究对象的多种属性和行为,运用各领域中的成熟理论进行分析有助于把握研究对象的总体状况,为实际问题的解决提供依据。为写作本书,笔者进行了大量的文献研究和实地调研,对理论研究中的理想状态与实际状况的差距有深刻体会,认识到理论必须为实践服务,实践活动应当通过理论进行提升,避免单纯的思辩,防止陷入实践细节的泥沼。

附录 1 参考文献

1. 中华人民共和国档案法. 北京: 档案出版社, 1997.
2. 中华人民共和国档案法实施办法. 北京: 档案出版社, 1999.
3. 国家档案局. 档案工作基本术语. 北京: 档案出版社, 1996.
4. 国家档案局. 档案工作基本术语. 北京: 档案出版社, 1996.
5. 国家档案局. 档案工作基本术语. 北京: 档案出版社, 1996.
6. 国家档案局. 档案工作基本术语. 北京: 档案出版社, 1996.
7. 国家档案局. 档案工作基本术语. 北京: 档案出版社, 1996.
8. 国家档案局. 档案工作基本术语. 北京: 档案出版社, 1996.
9. 国家档案局. 档案工作基本术语. 北京: 档案出版社, 1996.
10. 国家档案局. 档案工作基本术语. 北京: 档案出版社, 1996.

第2章 档案 数据库概述

档案数据库是档案信息资源建设的主流组织形式，研究档案数据库必须从档案和数据库管理的角度进行综合审视，梳理、界定其基本概念，了解其基本特点和局限，明确其表现形式和基本类型，把握其现状和意义，熟悉研究对象各方面属性的表现。

2.1 档案信息资源组织的三个阶段

作为信息活动的重要组成部分，信息组织活动一直与人类社会的产生发展紧密联系在一起。早在原始社会，我们的祖先就通过“结绳记事”将实践中获得的信息组织记录在绳结上，以此来记录氏族活动。随着人类社会的不断延续，信息包括档案信息的组织形式也一直在不断发展。笔者不揣冒昧，将档案信息组织技术的发展进行了粗糙地划分，以期描述档案信息组织技术的发展轨迹，认识数据库这种组织形式在档案信息组织管理技术中的地位。

2.1.1 偏重于记录技术的早期档案管理

档案作为人类重要的记忆资源，不同时代的人们都利用各自历史条件下的最佳技术手段来进行记录和**维护**。档案在我国进行记录的载体形式，历经甲骨金石、竹简木牍、**缣帛漆牍**、**金册铁券**，以至于纸墨文书。早在公元前三千年的殷商时期，我国的先人们就在龟甲兽骨上刻写文字，这些记录王室占卜事宜的甲骨时隔几千年后重新被人们发现，刻写的信息仍然能够被破译、识读，使我们可以穿越历史时空了解当时的社会情况；殷商周时代还遗留下刻有铭文的青铜器；春秋时期制作的金文档案记录了许多重要事件和法令，时至今日仍令人叹为观止。此外，历史上，还有大量的文献书写在竹片（简）、木版（牍）、丝织品（缣帛）之上，相对金石而言，这些载体的书写更为便捷，传输更为方便。笔者将纸张载体出现以前的档案管理界定为早期档案管理阶段。

通过对这些信息记录形式的简单考察，笔者认为早期档案侧重档案记录技术的发展。像甲骨金石，它们面向祭祀和法律，重点在表现权威，需要坚硬、耐久性好、具有一定象征意义的载体来进行记录。此外，对于一般性质的文书，古人则一直在寻找最佳的记录手段，直至纸张的发明。在此过程中，针对形态各异的载体，发展了不同的物理组织方式，但总的来说，由于受到载体形态的影响，这个时期的档案文献多是“单篇流传”，除非是由官方组织进行的有一定规模的编修，个体很难进行有效的信息组织活动。

早期档案的信息组织形式相对都比较简单。比如，甲骨的物理组织方式就非常简朴，通常是集中地保存于宗庙所在地由官方管理，并辅以简单排序，如按朝代、按龟甲和牛胛骨分别保存，有的龟甲还专门穿洞以便编制成册。而金文档案由于形制特别，还未见有关于其信息组织形式的论述。简牍缣帛档案载体相对轻便，信息的物理组织也有了较为灵活的形式。人们还通过将竹简连编，