

数字档案

赵豪迈 著

长期保存研究



陕西师范大学出版总社

赵豪迈 著

数字档案 长期保存研究



陕西师范大学出版总社

图书代号: ZZ15N0263

图书在版编目(CIP)数据

数字档案长期保存研究 / 赵豪迈著. —西安: 陕西师范大学出版总社有限公司, 2015. 3
ISBN 978-7-5613-8102-1

I. ①数… II. ①赵… III. ①数字技术—应用—档案管理—研究 IV. ①G270.7

中国版本图书馆 CIP 数据核字(2015)第 061006 号

SHUZI DANGAN CHANGQI BAOCUN YANJIU

数字档案长期保存研究

赵豪迈 著

责任编辑 /	王红凯
责任校对 /	梁菲 华翔凤
封面设计 /	鼎新设计
出版发行 /	陕西师范大学出版总社 (西安市长安南路 199 号 邮编 710062)
网 址 /	http://www.snupg.com
经 销 /	新华书店
印 刷 /	北京京华虎彩印刷有限公司
开 本 /	880mm × 1230mm 1/32
印 张 /	8.25
插 页 /	2
字 数 /	180 千
版 次 /	2015 年 3 月第 1 版
印 次 /	2015 年 3 月第 1 次印刷
书 号 /	ISBN 978-7-5613-8102-1
定 价 /	20.00 元

读者购书、书店添货或发现印刷装订问题,请与本社高教出版分社联系调换。
电 话:(029)85303622(传真) 85307826

摘 要

相对于一般电子文件的长期保存而言,本书研究并解决的是基于网络环境下的数字档案长期保存机制,因其使用环境的复杂性导致对其进行的研究必将涉及数字档案传播的整个生命周期,即数字档案的形成阶段、保存阶段和利用阶段。形成阶段是基础,保存阶段是关键,利用阶段是根本。本课题围绕数字档案生命周期内每个阶段的任务分别展开,以期通过对三者的有效整合为数字档案的长期利用提供一整套基于网络环境下符合标准的数字档案长期保存系统框架。

数字档案是随着计算机网络技术、数据库技术以及多媒体技术的发展而产生的一种新型档案信息形态,具有信息的可变性、载体的脆弱性、信息与载体的可分离性、对系统的依赖性、内容传输的网络性、信息组织形式网状化等特点。数字档案与纸质档案相比,正处于“成长”的危机之中,随时都有可能因某种技术、经济、政治或法律等原因使这种危机爆发为获取数字档案信息的严重障碍。由于数字档案尚处于起步阶段,人类数字化的知识、文化成果应得到长期保存还未被充分认识,人们的数字档案归档意识还比较淡薄,因此数字档案管理目前存在着诸多难点和问题。

随着数字档案数量日益增多,种类日益丰富,重要性逐渐增强,如何构建合理的数字档案长期保存体系,制定合理的数字档案长期保存策略,将有价值的数字档案进行长久保存,确保各种数字档案能随时提取利用,是档案馆当前亟待解决的问题。目前我国各级档案馆的数字档案长期保存情况不容乐观,存在数字档案长期保存战略目标不清晰、长期保存理论和实践研究不深入、长期保存的法规制度不健全等问题。但由于上述情况存在的原因是多层面的,受载体因素、技术因素、法律因素、经济因素等整体环境的影响,因此,应当为我国的数字档案资源保存系统构建一个数字档案资源长期保存的国家策略。

电子文件是数字档案的生命之源,基于文件生命周期理论,应当从形成阶段开始就对电子文件的真实可靠等特性进行维护,以保证电子文件能够顺利进入长久保存阶段,这种前端控制思想对电子文件的长久保存起着至关重要的作用。电子政务环境下,电子文件的形成过程指的是文件经过起草、修改、审核、签发等一系列环节最终形成具有使用价值的文件信息的过程,各个环节对文件的形成都有着不可或缺的重要作用,且哪个环节出现差错都会对文件的长久保存构成威胁。因此,必须通过对电子文件的前端控制来减少电子文件长久保存的后期管理风险,并通过对电子文件形成阶段的风险识别及关键性风险的分析,依托管理和技术手段初步构建一个覆盖内、外部领域的,全方位的、基于形成阶段的风险应对体系。

经过长期的技术探索,目前形成了迁移技术、更新技术、转换技术、仿真技术等多种数字档案长期保存技术。尽管如此,目前

数字档案的长期保存仍旧遇到了前所未有的挑战,成为整个数字档案管理工作的重点与难点。由于数字档案长期保存技术的复杂性和动态变换的特点,因此很难建立一个完整规范的技术体系,但是我们确实有必要把现有的技术进行适当的梳理、归纳,并对这些技术进行分析,形成数字档案长期保存技术选择策略。数字档案迁移技术成熟,风险小,操作简便,成本低,是目前数字档案长期保存的主要技术策略。另外,笔者对音视频数字档案长期保存处理技术也进行了初步研究。

数字档案文件保存格式种类繁多,不同类型的数字档案有不同种类、不同机构推出的存储格式。一般而言,存储格式应当选择符合国家标准格式,暂时未制定标准的,应选择开放格式或主流格式。常见的适用于数字档案长期保存的文件格式有 PDF/A、TIFF、JPEG2000、FLAC、MPEG 等。数字档案长期保存文件格式的选择是由电子文件创建过程的操作要求和数字档案长期存取的要求所决定的,其选择的标准应遵循开放性、标准性、稳定性、广采性、软硬件平台的独立性、真实性、可利用性等原则,以确保数字档案长期可读、可利用。数字化音频文件的存储格式最好选择 CD 或 MP3 格式,数字化视频文件推荐使用 MPEG 格式。数字档案长期保存技术可分为在线、近线、离线、异地与异质等存储方法,实际应用中一般选择几种存储方法相结合的方式分级存储,既确保安全高效,又兼顾经济实用。数字档案主要的存储介质为光盘、磁带和磁盘。选择数字档案存储介质时需要考虑多种因素,如相关的国际、国家标准或推荐意见,数字档案的生命周期,存储媒体的自身因素,部门的具体情况等。存储介质的选择

标准不是绝对的,档案机构在选择存储介质时,既要考虑符合标准的数据保存要求,也要权衡自身的经济承受能力。光盘存储技术具有高密度海量存储、功能较完备的信息处理技术、较长的保存寿命、较低的成本以及使用便利等优点,是当前数字档案长期存储的主要手段。但我们应结合光盘存储技术规范应用数字档案光盘存储技术,最大限度地保护光盘,延长其寿命。数字档案存放处的状况对电子载体及电子文件信息安全的影响也是不言而喻的,创造并维护一个符合数字档案保存条件的环境无疑有利于数字档案的信息安全性和保存的长久性。数字档案长期保存库房应符合温湿度适中、防止强磁场干扰、防火、防尘、防水等方面的要求,管理人员还应建立严格的数字档案库房管理制度。

数字档案在归档前,必须经过鉴定,鉴定在整个归档工作中占据很重要的地位。数字档案的归档范围,应参照执行国家档案局关于《机关文件材料归档和不归档的范围》的规定和其他有关科技文件、专门文件归档范围的规定,并结合数字档案的特点将反映机构主要活动、具有查考利用价值的电子文件纳入归档范围。数字档案的归档应结合数字档案元数据管理规范进行。数字档案可以借鉴国内外数字信息的呈缴制度以及登记备份制度,以更好地长期保存数字档案。必须建立数字档案的安全保障机制,加强数字档案的信息安全和数据库安全。在数字档案长期保存中,管理应从整体上进行宏观规划、组织和控制,这是数字档案长期保存的基础与保障。必须建立数字档案长期保存的管理机制,并基于文件生命周期理论和层次管理理论,建立起数字档案长期保存的管理规范和数字档案管理系统建设规范。

数字档案长期保存技术利用体系是一个复杂的有机组合体。OAIS 参考模型虽然从总体上对数字保存利用系统进行了考虑,并提供了一个概念框架,但由于数字档案长期保存技术的复杂性和动态变换的特点,因此很难建立一个完整规范的技术体系。有必要把现有的技术可能性进行适当的梳理归纳,提出一个可供参考的技术体系,并对这些技术体系之间的关系和作用机制进行分析,形成数字档案长期保存机制。与其他数字信息保存系统相比,数字档案长期保存利用机制复杂,涉及数据摄入、档案存储、数据管理、数据分发、内容管理、数据库系统、元数据管理、搜索引擎、分布式文件系统、唯一标识系统、存储系统、存储管理系统、媒体检测(监测)系统、认证授权系统、安全机制、统计管理等多个领域,覆盖多方面技术,构成一个复杂的技术体系。可以看到,数字档案长期保存利用技术分属于保存管理、摄入管理、仓储管理、 workflow 管理和访问管理五个功能块,这五个功能块各司其职并相互作用,形成基于事务的工作流触发机制的数字档案长期保存利用技术策略。

关键词:数字档案;长期保存;机制

目 录

第一章	数字档案的概念、内涵与特点	(1)
1.1	数字档案的概念和内涵	(1)
1.2	数字档案的特点	(3)
1.3	数字档案的管理难题	(4)
1.3.1	数字档案长期保存的价值问题	(4)
1.3.2	数字档案的长期存取问题	(6)
1.3.3	数字档案数字化后的失真问题	(7)
1.3.4	数字档案管理中的安全问题	(8)
1.3.5	数字档案管理中的诸多技术问题	(8)
1.3.6	数字档案长期保存的标准化问题	(9)
1.3.7	数字档案管理中的法规问题	(10)
第二章	数字档案长期保存的现状与问题	(12)
2.1	数字档案长期保存的概述	(12)
2.2	数字档案长期保存的现状	(14)
2.2.1	数字档案长期保存的战略目标不清晰	(15)
2.2.2	数字档案长期保存的理论和实践研究不深入	(15)
2.2.3	数字档案长期保存的法规制度不健全	(16)
2.3	数字档案长期保存的现状分析	(17)

2.3.1	内因分析	(17)
2.3.2	外因分析	(18)
2.4	数字档案长期保存需要解决的问题	(20)
2.4.1	资源的选择问题	(21)
2.4.2	技术保障机制	(22)
2.4.3	标准	(23)
2.4.4	管理方面	(24)
2.5	数字档案长期保存的未来展望	(24)
第三章	数字档案的形成阶段与长期保存	(26)
3.1	电子文件的形成阶段与长久保存	(26)
3.1.1	电子文件的形成过程与形成阶段	(28)
3.1.2	电子文件的形成阶段与其长久保存的关系	(36)
3.2	电子文件形成阶段的风险分析	(41)
3.2.1	电子文件形成阶段的风险识别模型	(41)
3.2.2	电子文件形成阶段长久保存的质量目标	(43)
3.2.3	电子文件形成过程中的关键风险因素	(44)
3.2.4	电子文件形成阶段的风险因素对其长久保 存的影响	(47)
3.3	电子文件形成阶段的风险应对策略	(49)
3.3.1	电子文件长久保存的风险应对体系	(50)
3.3.2	外部领域电子文件风险的应对措施	(55)
3.3.3	内部领域电子文件风险的应对措施	(57)

第四章	数字档案长期保存的技术与选择	(64)
4.1	数字档案长期保存技术	(64)
4.1.1	迁移技术	(65)
4.1.2	更新技术	(67)
4.1.3	转换技术	(68)
4.1.4	仿真技术	(69)
4.1.5	数字图形输入板技术	(70)
4.1.6	再生性保护技术	(71)
4.1.7	计算机软硬件技术档案馆	(72)
4.1.8	封装技术	(72)
4.1.9	网络存储技术	(73)
4.1.10	管理工具	(75)
4.2	数字档案长期保存技术的选择策略	(76)
4.2.1	数字档案长期保存技术的策略分析框架	(78)
4.2.2	数字档案长期保存的层次划分	(79)
4.2.3	数字档案长期保存技术的策略分析	(81)
4.3	数字档案迁移操作实务	(92)
4.3.1	数字档案的迁移方式	(92)
4.3.2	数字档案迁移的传输技术	(94)
4.3.3	数字档案迁移中的安全保障措施	(96)
4.4	音视频数字档案长期保存的处理技术	(98)
4.4.1	音频数字档案长期保存的处理技术	(98)
4.4.2	视频数字档案长期保存的处理技术	(99)

第五章	数字档案长期保存的格式与载体	(101)
5.1	数字档案长期保存的文件格式	(102)
5.1.1	数字档案长期保存的常见文件格式	(103)
5.1.2	数字档案长期保存的文件格式选择	(106)
5.1.3	音视频数字档案长期保存的文件格式选择	(108)
5.2	数字档案长期保存的载体和存储系统	(111)
5.2.1	数字档案长期保存的技术方法	(111)
5.2.2	数字档案长期存储介质的选择	(112)
5.2.3	数字档案光盘存储技术规范	(116)
5.2.4	数字档案光盘存储技术机制	(124)
5.2.5	数字档案长期存储环境的保护	(132)
第六章	数字档案长期保存的管理机制	(136)
6.1	数字档案的鉴定与归档	(137)
6.1.1	数字档案的鉴定	(137)
6.1.2	数字档案的归档	(138)
6.1.3	数字档案的元数据管理	(140)
6.2	数字档案的呈缴与登记备份	(146)
6.2.1	数字档案的呈缴	(146)
6.2.2	数字档案的登记备份	(153)
6.3	数字档案的安全保障	(164)
6.3.1	数字档案的信息安全	(164)
6.3.2	数字档案的数据库安全	(165)
6.4	数字档案的管理规范	(169)

6.4.1	数字档案的管理机制	(169)
6.4.2	网络环境下的电子文件管理规范	(170)
6.4.3	数字档案管理规范	(177)
6.4.4	数字档案管理系统建设规范	(194)
第七章	数字档案长期保存的利用机制	(209)
7.1	数字档案长期保存的利用体系	(210)
7.1.1	数字档案长期保存的利用前沿	(210)
7.1.2	Kenneth Thibodeau 的二维坐标模型	(221)
7.1.3	NARA 的 EAR 模型	(223)
7.1.4	Priscilla Caplan 的金字塔模型	(225)
7.1.5	数字档案长期保存的利用技术体系	(226)
7.2	数字档案长期保存机制	(228)
7.2.1	长期保存功能	(229)
7.2.2	摄入管理功能	(231)
7.2.3	仓储管理功能	(233)
7.2.4	workflow 管理功能	(237)
7.2.5	访问管理功能	(239)
参考文献		(241)
后记		(250)

第一章 数字档案的概念、内涵与特点

本章摘要

数字档案是随着计算机网络技术、数据库技术以及多媒体技术的发展而产生的一种新型档案信息形态,具有信息的可变性、载体的脆弱性、信息与载体的可分离性、对系统的依赖性、内容传输的网络性、信息组织形式网状化等特点。数字档案与纸质档案相比,正处于“成长”的危机之中,随时都有可能因某种技术、经济、政治或法律等原因使这种危机爆发为获取数字档案信息的严重障碍。由于数字档案尚处于起步发展阶段,人类数字化的知识、文化成果应得到长期保存还未被充分认识,人们的数字档案归档意识还比较淡薄,因此数字档案管理目前存在着诸多的难点和问题。

1.1 数字档案的概念和内涵

数字档案是随着计算机网络技术、数据库技术以及多媒体技术的发展而产生的一种新型档案信息形态。数字档案以数字化的形式存储,网络化的形式获取,并利用计算机系统进行管理,形

成一个结构有序的数字档案信息库,及时提供利用,实现数字档案信息资源共享。

目前档案界对数字档案概念的认识并不统一和清晰。其一,对数字档案概念的研究并未引起档案界太多的深入关注,流于概念的多,深入探究的少;其二,一些对数字档案概念的认识,大部分都是以传统的档案概念为基础的,仍旧把数字档案当作电子文件的转化物来认识,无非是传统档案概念在数字时代的翻版。这种认识存在一定的模糊性。在数字时代的信息网络环境下,电子文件、数字档案、电子档案、数字情报、数字资源、数字文献等有什么不同?如何区分它们?如果不能把数字档案从其他这些数字信息中区分出来,那么数字档案概念的存在就成为问题。因此,必须把数字档案概念从数字信息中清晰地区别出来,才能使它成为一个独立发展、有其清晰内涵和外延的数字概念。

笔者认为,数字档案(Digital Archive)是存储于电磁介质上的,以0、1二进制编码的形式存在的,可以通过计算机网络传输的档案,包括利用计算机直接形成的档案和经由数字化之后转变为计算机和多媒体技术支持的纸质档案信息。

这个概念强调了两个方面的内容:一方面,强调了档案以数字形式存储和利用;另一方面,强调了它存储的信息是以电磁介质为载体,按0和1二进制编码的方法加以存储和处理,把原来用纸质存储的信息转变为用计算机和多媒体技术存储和处理的信息。

目前,数字档案信息的来源主要有两种途径,一是计算机直接生成的数字化档案信息,如通过网络获取的电子文件、电子政

务归档电子文件,办公现代化建设中直接形成的数字文件、数据库、音视频多媒体文件、程序软件及其目录数据库等;二是把大量现已存在的以不同形式和载体存储的信息资料,如纸质档案、图片档案、声像档案等,转化为以数字形式存储起来的数字化档案信息。

1.2 数字档案的特点

数字档案具有数字档案信息的可变性、数字档案载体的脆弱性、数字档案信息与载体的可分离性、数字档案对系统的依赖性、数字档案内容传输的网络性以及数字档案信息组织形式网状化等特点。同时,呈现出以下四大特点。

(1) 档案信息资源的数字化

数字档案资源主要包括由原来用纸质形式存储的信息转换成的计算机中的数字化信息,其中除大量的文字档案外,还有一些图形档案、图像档案以及音频、视频、动画、软件程序等电子档案。人们通过多媒体技术把多种档案形式有机地结合在一起进行统一存储和管理。

(2) 档案内容传输的网络化

网络和通讯系统是数字档案的重要基础,如果没有网络系统或网络系统不完善,数字档案就失去了它的真正意义。信息的网络传输使数字档案超越时空观念,跨越馆(室)藏档案的地域界限。