

数字档案馆多平台档案数据同步管理研究

前 言

在数字时代，档案资源的数字化、管理现代化和服务网络化是档案工作发展的战略方向。研究解决数字档案长久保存问题，使其能够在尽可能长的时间内保持真实性、完整性以及可用性，是全球档案界面临的共同挑战。2009年，云南省档案馆启动数字档案馆项目建设，以整合数字档案信息资源，有效管理归档电子文件；完成云南省档案馆馆藏档案数字化，强化档案安全管理；永久安全保存和开发利用数字化档案信息资源，提高档案工作为党委、政府和社会公众服务的能力为目标，建设基于 OAIS 模型的云南省省级数字档案管理平台，研究和实践数字档案资源的长期保存和便捷利用的科学方法和有效途径。数字档案馆多平台数据同步就是围绕这一目标实现开展的研究课题，目的是解决数字档案馆多平台部署环境下数字档案数据的一致性问题。

1 研究背景

1.1 OAIS 模型的形成与发展

为了解决数字信息长久保存问题，使其能够在尽可能长的时间内保持真实性、完整性以及可用性。1995年，美国国家航空和宇宙航行局（NASA）的空间数据系统咨询委员会（The Consultative Committee for Space Data Systems，简称CCSDS）着手开发“开放性档案信息系统参考模型（the Reference Model for an Open Archival Information System，简称OAIS）”，其目的在于对数字信息的存取和长期保存规定概念和参考框架，统一的数字归档系统基本功能需求以及统一的数字信息保存标准，用于指导人们建设统一的数字信息保存平台，实现数字信息的长久保存。2003年，OAIS参考模型于作为ISO的标准（ISO14721:2003）正式颁发。2012年，该标准进行了首次修订更新（ISO14721:2012）。

OAIS模型的提出为数字信息的长期保存提供以一种可行的方式，该模型描述了所有的归档信息保存功能，包括摄取、归档存储、数据管理、访问和分发等功能，提出了基于信息包的数字信息的长期保存方式。

OAIS参考模型主要包括功能模型和信息模型，前者为数字档案馆提供了良好的参考架构，能够保证系统架构的科学性，避免方向性错误。后者为数字档案馆各类信息规范，包括电子文件元数据规范与长久保存元数据的制定提供重

要的参考模板。

1.2 OAIS 模型在我国数字档案馆建设中的应用

OAIS 模型适用于所有致力于长期保存数字资源并提供利用的系统和组织,因此受到了包括档案机构在内的相关机构的广泛关注。目前,在我国档案领域 OAIS 模型作为数字档案信息长期保存和利用的基本框架体系广泛被应用在电子文件管理、元数据标准开发、数字档案资源的长期保存、数字档案资源管理系统开发、数字档案馆建设等方面,在理论研究、标准制定和应用实践方面取得了丰富的成果。国家档案局制定《数字档案馆指南》以 OAIS 模型为基本框架,大量借鉴 OAIS 的成果;档案行业标准 DAT/47-2009《版式电子文件长期保存格式需求》、DAT/48-2009《基于 XML 的电子文件封装规范》对 OAIS 标准进行了引用。国家档案局技术部组织研发了数字档案移交接收长期保存系统,江西省档案馆、中科院档案馆、云南省档案馆等多个档案机构建设了基于 OAIS 模型的数字档案馆,积累大量的经验。应用 OAIS 模型指导数字档案的长期保存,建设数字档案管理系统、数字档案馆系统,已经逐步成为我国档案领域的共识。

1.3 OAIS 模型在云南省档案馆数字档案馆建设中的实践

2009 年,云南省档案馆启动数字档案馆建设,并被国家档案局确定为《数字档案馆建设指南》应用的省级试点单位。根据项目规划,云南省档案馆数字档案馆建设计划投资 9164 万元,完成云南省档案馆馆藏档案目录 1400 万条和 1

亿 4000 万页原文的数字化处理；完成云南省档案馆馆藏录像、录音、实物和缩微胶片等 1000 余万件多媒体资料的数字化处理；建成面积为 1200 平方米的多媒体展示平台；建成面积为 577 平方米、能容纳 120 人的电子档案查阅场所；实现与云南省 130 余个省级机关电子档案归档和查阅的网络化应用；面向社会公众提供档案信息资源检索服务。项目建设周期 10 年，分两个阶段实施。第一阶段 2009 年至 2012 年，投资 3264 万元，主要完成数字档案馆的基础业务平台建设和部分馆藏档案数字化扫描；第二阶段 2013 年至 2018 年，投资 5900 万元，重点是馆藏档案数字化扫描和基础平台升级完善。

云南省档案馆数字档案馆的规划、设计和建设以《数字档案馆建设指南》为指导，广泛借鉴和吸收了国内江西、山东、广东、青岛、绍兴等省、市数字档案馆建设经验，紧密结合自身的业务需求和实际情况，遵循 OAIS 模型，设计了“一库四平台三体系”多平台部署的系统架构，实现了基于 OAIS 模型的业务框架和系统功能，成为 OAIS 模型应用的又一实例。

1.3.1 “一库四平台三体系”多平台部署的系统架构

系统建设内容概括起来为“一库四平台三个体系”，包括：一是建设全面覆盖馆藏、内容丰富的云南省档案信息资源数据库；二是建设数字档案信息资源管理平台、政务信息化档案资源管理平台、多媒体档案采编及展示平台和档案信

息资源共享服务平台,面向政府机构和社会公众提供方便快捷、安全可靠的档案信息资源网络化服务;三是建立基础支撑体系、电子档案安全体系和标准规范体系建设,保障档案信息资源数据库和应用平台的安全稳定运行;四是电子档案查阅场所建设;五是档案信息化专业技术队伍建设。

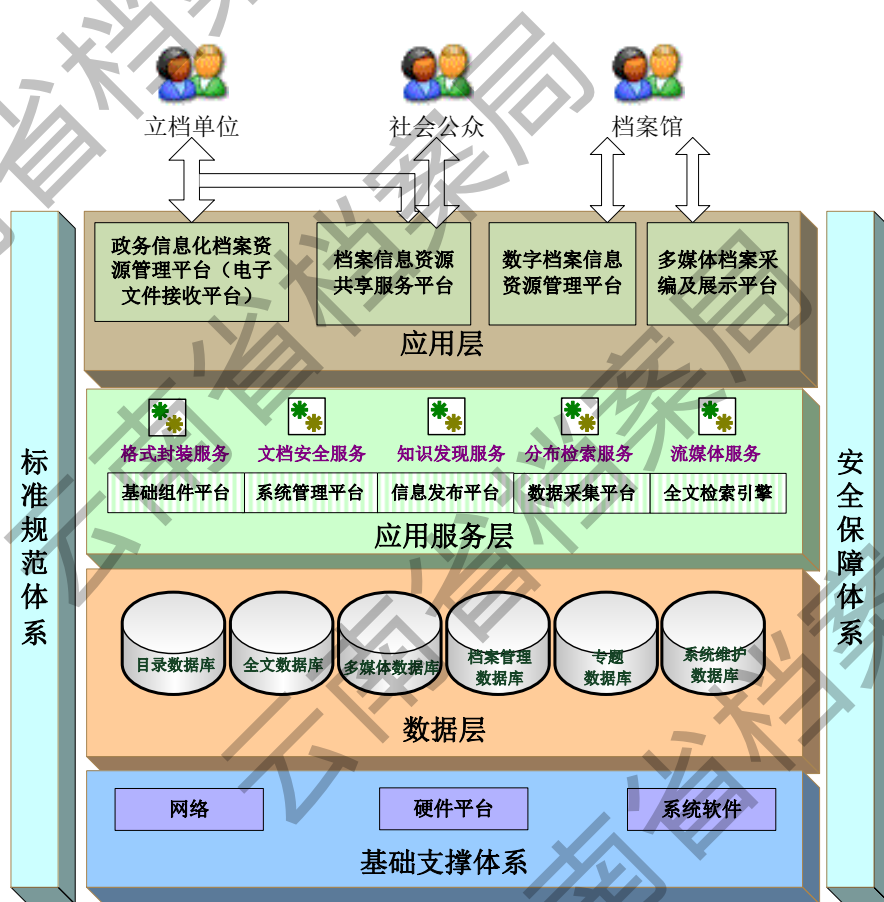


图 1 云南省档案馆数字档案建设系统体系架构图

基础支撑体系：包括网络平台、硬件平台和系统软件平台以及各种网络协议(HTTP/SSL/FTP/SMTP/ SOAP /MMS , 分别为超文本传输、传输加密、大文件传输、数据交换、流媒体播放), 是系统得以部署和运行的基础条件。

数据层 :即档案信息资源库 ,是数字档案馆系统的核心。

按照资源的类型和用途可以分成目录数据库、全文数据库、多媒体数据库、档案管理数据库、专题数据库和系统维护数据库。

应用服务层：指应用基础平台和第三方工具软件，包括基础组件平台、系统管理平台、信息发布平台、数据采集平台、全文检索引擎、格式转换及封装服务等，为应用系统的建设提供强有力的平台支撑。

应用层：建成四个应用平台，包括数字档案信息资源管理平台、政务信息化档案资源管理平台、多媒体档案采编及展示平台、档案信息资源共享服务平台。

安全保障体系：包括网络安全、系统安全、数据库安全、信息安全、设备安全、信息介质安全和病毒防犯等各个方面。

标准规范体系：系统建设的同时，需要建立相关的技术管理规范 and 标准，如电子文件格式标准、元数据标准等。

根据档案信息及其服务对象的安全保密保护等级的不同，为了确保档案数据和系统的安全，按照信息系统安全保密管理有关要求，数字档案馆所包含的数字化档案信息资源管理平台、政务信息化档案资源管理平台、多媒体档案采编及展示平台、档案信息资源共享服务平台等应用平台在相互物理隔离的政务网、局域网、互联网、核心网四个网络平台上进行部署，形成了多平台部署的数字档案馆系统。

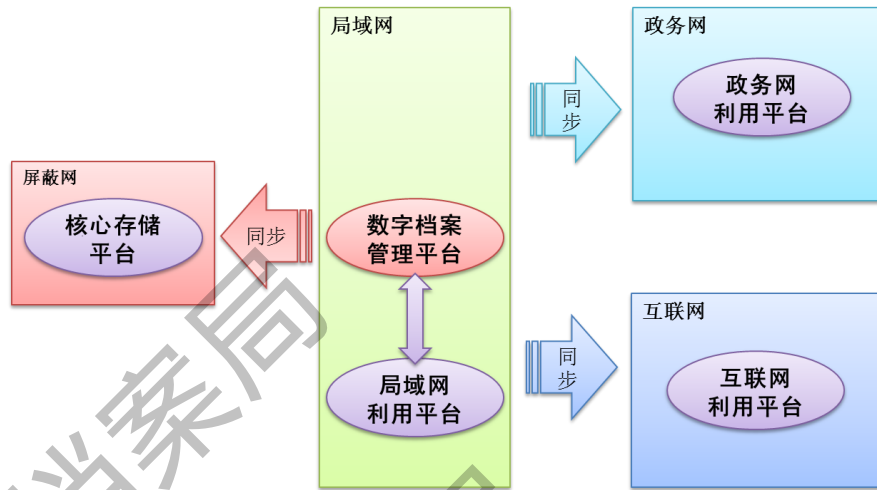


图 2 云南省档案馆数字档案馆系统多平台部署示意图

1.3.2 基于 OAIS 模型的业务应用框架

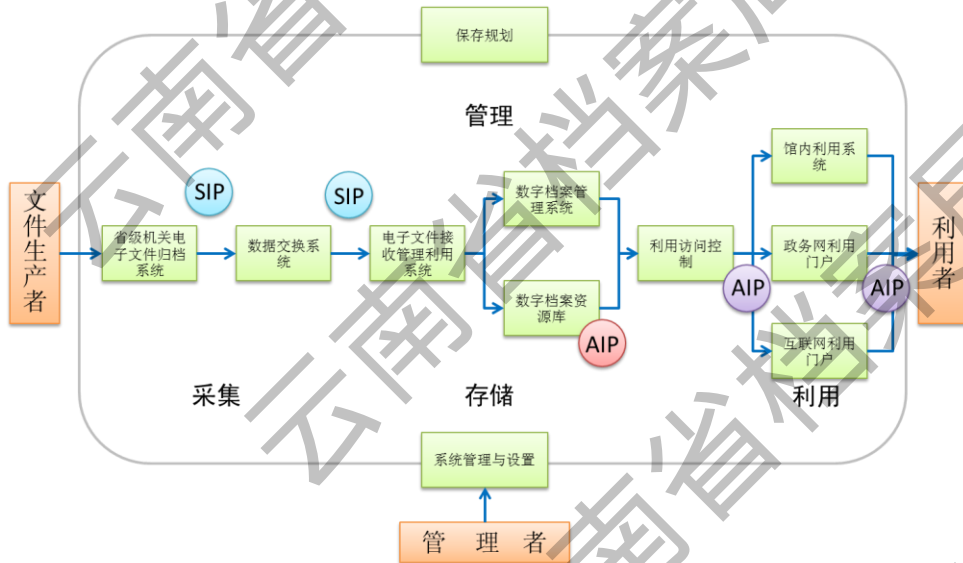


图 3 云南省档案馆数字档案馆 OAIS 模型应用框架图

云南省档案馆数字档案馆业务应用框架涵盖 OAIS 功能模型的 6 个功能模块包括采集、保存规划、管理、存储、访问、系统管理等：

1.3.2.1 采集

通过省级机关电子文件归档系统采集来自省直机关、省级业务系统的电子档案数据,通过数字化采集来自馆藏档案的数字化档案数据,并将采集的档案数据按省档案馆的归档要求进行有效整合。

1.3.2.2 保存规划

制定电子档案长期保存策略可作为本系统建设的配套规范。实现 SIP、AIP、DIP 三类信息包的定义和封装。

1.3.2.3 管理

档案馆数字档案的接收、挂接、录入、导入、征集、组卷、著录、鉴定、库房、编研、统计。

1.3.2.4 存储

存储管理、备份恢复、保存封装。

1.3.2.5 利用

档案查询、借阅利用、网站发布。



图 4 数字档案馆局域网利用平台



图 5 数字档案馆政务网利用平台



图 6 数字档案馆互联网利用平台

1.3.2.6 系统管控

日志管理、编目管理、数据建模、统一身份认证、用户权限管理、流程建模、全宗管理、界面建模。

1.3.2.7 信息包

根据 OAIS 模型，在数字信息的收、管、用三个不同阶段分别定义了 SIP、AIP、DIP 三类信息包。在云南省数字档案馆的业务应用框架中，根据国家档案局《DAT 46-2009 文书类电子文件元数据方案》、《DAT/48-2009 基于 XML 的电子文件封装规范》规定的元数据方案、对象封装标准，实

现了三类 EEP 格式信息包的封装，并按照收、管、用三个不同阶段的信息特征进行了不同的动态数据打包、拆包、封装的信息处理，并动态地贯穿于六大功能模块的应用及操作之中。

2 研究意义

构建基于 OAIS 模型的数字档案馆，是当前各国档案界解决数字化档案资源长期永久保存问题的重要实践。OAIS 模型作为数字信息长期保存的系统参考模型，为数字档案馆建设提供了基本框架，是数字档案馆系统建设的起点。如何以 OAIS 模型为起点，开展在不同环境中的应用实践，不断丰富和完善 OAIS 模型，提高其适应性，是 OAIS 研究和应用的重要内容。本课题的提出就是为了研究解决基于 OAIS 模型的数字档案馆系统(本课题研究的数字档案馆系统限定为基于 OAIS 模型的构架的数字档案馆系统，以下简称为数字档案馆系统)多平台部署环境下档案数据的一致性问题。其研究的意义集中体现在：

2.1 是构建和完善多平台环境下数字档案馆系统档案数据管理流程，细化和拓展 OAIS 模型档案应用的需要。

OAIS 为数字档案信息长期保存工作的开展和系统构架提供了指南和框架，但并没有局限于任何特定计算机平台、系统环境、系统设计范例、系统开发方法、数据库管理系统、

数据库设计范例、数据定义语言、命令语言、系统界面、用户界面、技术、所需媒体等系统建设要素。因此，也没有提供再多平台物理隔离网络环境下部署和应用的具体流程，因此，需要以 OAIS 参考模型为起点，综合物理隔离的网络环境特点、档案管理业务的具体需求和数字档案管理要求进行研究和分析，梳理和区分各构成模块的业务功能、业务流程数据流向和安全边界，提出多平台物理隔离环境下数字档案馆系统具体应用模型。

2.2是确保多平台环境下数字档案馆系统的档案数据一致性需要。

在数字档案馆系统中，对数字档案的管理是涵盖接收、整理、存储和利用全过程管理，这是确保数字档案真实、完整和有效的基本要求。在多平台部署的应用环境下，数字档案的整理、存储一般在局域网络平台上进行，数字档案的利用分别在局域网、政务网、互联网平台上实施，网络平台的物理隔离势必隔断数据管理的完整流程，导致管理流程截断和管理的 inconsistence。同时，也会导致同一档案数据在不同平台呈现不同的内容，从而影响数字档案的真实、完整和有效。因此，有必要研究和设计物理隔离多平台环境如何建立有效和可靠的数据同步机制，以实现管理和数据的一致性，确保档案数据的真实、完整和可用。

2.3是确保多平台部署环境下数字档案馆系统档案数据安全的需要。

为了确保计算机信息系统及信息的安全,我国相关法律制度对网络和信息的隔离做了明确规定,不同安全保密等级的数据必须在具备相应能力的网络平台上处理,各网络之间应采取物理隔离措施。《计算机信息系统保密管理暂行规定》(国家保密局国保发(1998)1号):“第十一条:国家秘密信息不得在与国际网络联网的计算机信息系统中存储、处理、传递”。《涉及国家秘密的通信、办公自动化和计算机信息系统审批暂行办法》(国家保密局国保发(1998)6号):“第十六条:涉密系统不得直接或间接与国际联网,必须实行物理隔离”。《计算机信息系统国际联网保密管理规定》(国家保密局国保发(1999)10号):“第六条:涉及国家秘密的计算机信息系统不得直接或间接地与国际互联网或其他公共信息网连接,必须实行物理隔离”。国务院办公厅《关于进一步推进全国政府系统办公自动化建设和应用工作的通知》(国办发(2000)36号):“(四)全国政府办公业务资源网及各地区各部门的办公业务网,必须在物理上与国际互联网完全隔离”。中共中央办公厅、国务院办公厅《关于转发国家信息化领导小组〈关于我国电子政务建设指导意见〉通知》(中办发(2002)17号):“(一)建设和整合统一的电子政务网络。电子政务网络由政务内网和政务外网构成,两网之间物理隔离,政务外网与互联网之间逻辑隔离”。中共中央保密委员会《关于严禁用涉密计算机上国际互联网的通知》(中保委发(2003)4号):“一、涉密计算机信息系统必须

同互联网实行物理隔离 ,严禁使用处理国家秘密信息的计算机上互联网 , 违者严肃查处。” 。

数字档案馆系统是管理国家档案资源的计算机信息系统 ,其管理的对象和提供的服务涉及不同等级的安全保密要求 ,为确保安全 ,必须在具备相应安全保密能力的平台上进行管理、存储和利用 ,这些平台之间必须进行物理隔离。因此 ,数字档案馆的管理功能必须在相应的平台上部署和实现 ,涉及的数字档案资源就必须在相应的平台上进行存储、管理和利用。因此 ,需要研究适应档案数据和系统安全管理的技术和管理机制 ,解决物理隔离环境下数据管理的安全问题。

3 研究依据

本课题依据国家档案局《2012 年度国家档案局科技项目计划》(档函【2012】97 号) 经国家档案局立项 , 由云南省档案局承担研究任务。

4 研究目标

4.1 数据同步的目标

分析和提出数据同步所要达到目标 ;

4.2 同步数据的类型

归纳和梳理数字档案馆系统中需要进行多平台同步的数据类型和特点；

4.3 数据同步的管理要求

从法律和业务规范的角度，研究和分析数据同步的要求，包括档案原始性、真实性、完整性保证、安全控制等；

4.4 数据同步的技术要求

从技术的角度，研究和分析数据同步要求，包括数据同步的功能、性能、可靠性、可用性等要求；

4.5 数据同步的业务模型

基于数据同步目标、要求分析，提出数据同步的业务模型。

5 研究过程

5.1 文献综述

近年来，基于 OAIS 数字档案馆系统的建设是我国档案界理论研究和实践探索的重要内容，而物理隔离环境下信息系统建设与应用、数据同步、分布式数据库等通用信息技术的则更为丰富。这些成果是本课题的研究的实践基础和理论支撑。

5.1.1 文献分布

根据 2013 年 11 月课题组就数字档案馆、OAIS、数据同步、数据交换、分布式数据库、物理隔离等关键词对中国

知网 (www.cnki.com.cn) 中国学术期刊网络出版总库、中国博士学位论文全文数据库、中国优秀硕士学位论文全文数据库、中国重要会议论文全文数据库、中国重要报纸全文数据库进行主题检索，得的相关文献数据如下：

	数字档案馆	OAIS	数据同步	数据交换	分布式数据库	物理隔离
中国学术期刊网络出版总库	1981	113	15196	22834	5091	1406
中国博士学位论文全文数据库	3	6	2615	2330	401	232
中国优秀硕士学位论文全文数据库	124	22	12674	15058	5000	828
中国重要会议论文全文数据库	102	5	1026	1382	389	113
中国重要报纸全文数据库	200	8	149	370	43	234
合计	2410	154	31660	41974	10924	2813

表 1 单一关键词中国知网主题文献统计

对上述关键词组配后，进行主题检索，结果如下：

	数字档案馆 数据同步	数字档案馆 数据交换	数字档案馆 物理隔离	OAIS 数据同步	OAIS 数据交换	OAIS 物理隔离
中国学术期刊网络出版总库	3	7	1	0	1	0
中国博士学位论文全文数据库	0	0	0	0	0	0
中国优秀硕士学位论文全文数据库	1	4	1	0	0	0
中国重要会议论文全文数据库	0	0	0	0	0	0
中国重要报纸全文数据库	0	0	0	0	0	0
合计	4	11	2	0	1	0

表 2 组配关键词中国知网主题文献统计

5.1.2 文献分析

分析以上文献主题统计结果,与本课题相关的研究呈现以下特点:

5.1.2.1 数字档案馆及 OAIS 相关主题研究形成一定规模,但文献总量偏少,特别是缺乏高层次的研究文献。两个主题文献量合计 2564 篇,OAIS 主题文献量为 154 篇,还不到两项合计总量的 10%。博士层次的研究文献仅有 9 篇。

5.1.2.2 物理隔离、数据同步、数据交换、分布式数据库等相关通用信息技术的研究已经较为成熟,文献量大,高层次的研究文献已经形成一定规模。数据同步、数据交换两个主题文献量合集达到 73634 篇,分布式数据库主题文献量达到 10924 篇,博士层次的研究文献三个主题达到 5346 篇。

5.1.2.3 多平台数字档案馆数据同步研究基本处于空白状态。“数字档案馆”和“数据同步”组配检索结果文献为 4 篇,其中 1 篇为本课题的新闻报道外,其余 3 篇内容都与本课题研究主题没有直接关系。“数字档案馆”和“数据交换”组配检索结果文献为 7 篇,研究主题是数据交换的技术方法。“数字档案馆”和“物理隔离”组配检索结果文献为 2 篇,研究的主题是数字档案馆安全研究,与本课题的相关度较低。“OAIS”和“数据交换”、“OAIS”“数据同步”、“OAIS”“物理隔离”组配检索,只得到 1 篇文献。反映处基于 OAIS 模型数字档案馆系统多平台环境下数据同步综合性研究目前基本处于空白状态。

5.2 案例研究

以云南省档案馆数字档案馆系统为案例,梳理档案数据
和数据管理的情况,分析当前系统档案数据的类型和特点、
数据同步流程,按照档案及数据管理的原则、要求和标准,
对当前系统档案数据同步情况进行评价分析,总结经验,找
出问题。

5.3 实证研究

基于调查和分析的结果,归纳档案及数据管理的原则、
要求和标准,提出数字档案馆系统档案数据同步的管理和技
术要求,形成档案数据同步的业务模型。

5.4 研究过程

云南省档案局领受课题任务后,黄凤平局长组织召开课
题研究工作会,明确课题研究工作的人员安排、任务计划和
时间进度要求。课题组长张文芝召集课题组成员讨论研究思
路、方法、步骤和分工,确定课题研究的具体步骤:查阅现
有文献、学习相关制度规范,分析和掌握相关研究现状和已
有成果;调查云南省档案馆数字档案馆数据管理情况,梳理
总结实际工作经验和问题;分析数据同步的特点和规律,提
出数据同步的管理和技术要求,形成数据同步模型。从 2012
年 6 月份起,课题组成员按照分工逐项开展工作,先后查阅
大量文献、学习相关规范、开展实地调研,多次召开课题研
究分析会,统一观点和认识。2013 年 10 月底完成课题初稿
汇总,2013 年 12 月完成统稿。经过不断研究和完善,课题
达到预期研究目标,形成数据同步模型。2014 年 月,课题

组完成课题研究，上报国家档案局申请课题鉴定。2014年3月7日，国家档案局组织专家组对课题研究成果进行鉴定，与会专家通过听取汇报、查验文档、现场质询，一致认为“该项目研究技术路线清晰，研究方法正确，研究内容丰富”，同意课题通过鉴定。课题组会后根据专家提出的意见，对研究成果进行进一步的丰富和完善，重点针对鉴定意见中“进一步开展数字档案馆多平台档案数据同步管理研究成果的推广应用工作”的建议，结合云南省数字档案馆建设工作需要，将项目研究成果充分应用到数字档案馆建设管理实践中。2014年7月，云南省档案馆经过招标正式启动省数字档案馆系统软件升级改造项（招标编号：YZ2014083806147），以项目成果为基础开发建设云南省省数字档案馆档案数据同步功能模块，以期进一步提升省数字档案馆档案数据管理能力和水平，为省数字档案馆建设的长期可持续发展提供有力保障。

6 数字档案馆多平台数据同步的界定

6.1 相关概念

6.1.1 数据同步

数据同步是指数据在分布式应用环境下，使不同系统的数据之间或者数据的来源与其副本之间保持相互一致性。

6.1.2 数据同步的类型

根据同步通道不同，可分为在线同步和离线同步。在线同步是指系统之间有直接连接的网络通道，同步操作在系统之间直接进行。离线同步是指系统之间物理上是隔断的、无法直接通讯的，同步操作必须通过离线介质、网闸、数据渡船等中介间接进行。

根据同步对象不同，可分为数据库同步和文件同步。数据库同步一般是指基于数据库操作以数据库记录为单位进行的同步。文件同步一般是指基于文件系统操作以文件为单位进行的同步。

根据同步内容不同，可分为增量同步和差异同步。增量同步，是指来源系统和目标系统之间数据数量发生变化进行的同步，如来源系统的数据增加或者删除一条数据，目标系统的也要相应进行增加、删除，以保持数量一致。差异同步，是指来源系统和目标系统之间同一数据的来源和副本内容发生变化而进行同步。如来源数据某个字段的数值进行了更改，那么对应副本数据的字段值也要进行修改。

6.1.3 数据同步的过程

数据同步是对数据进行操作使数据副本与来源一致的过程。这个过程包括了变化捕捉、数据分发、数据更新、数据验证四个主要环节。

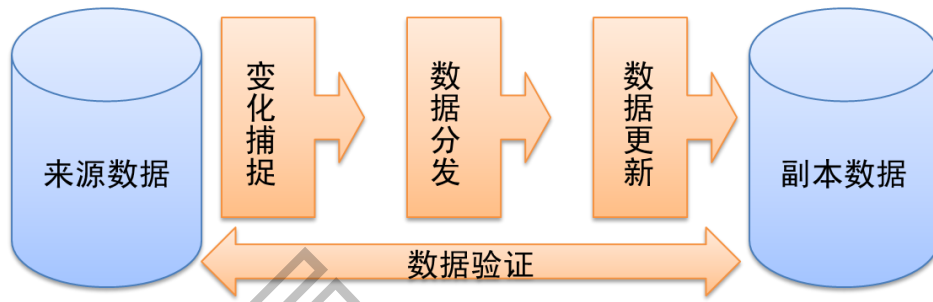


图7 数据同步一般过程

6.1.3.1 变化捕捉

发现和记录数据源的改变,这种改变导致数据源和副本不一致。因此,变化捕捉是数据同步的基础和起点。数据源的变化包含对数据的增加、删除和修改等。

6.1.3.2 数据分发

将已经变化的数据发送或者引起变化的事务(操作)发送到副本数据存放地。这种分发可以基于设定的规则自动进行,也可以人工进行启动。

6.1.3.3 数据更新

在副本数据存放地,以新的数据替换已有数据或者执行事务(操作),实现副本与源数据一致。

6.1.3.4 数据验证

在数据更新完成以后,比较更新过的数据与来源数据是否一致,一致则完成数据同步过程,不一致则进行出错处理。

6.2 多平台数据同步

多平台数据同步指在物理隔离的不同网络、系统的数据

库之间或者数据的来源与副本之间保持一致的过程。这里的平台包含系统、网络的两重意义，即应用系统（子系统）和物理网络（如政务网、互联网、局域网等）。同时，这些平台、网络之间进行了物理隔离。

6.3 数字档案馆系统多平台数据同步

数字档案馆系统多平台数据同步指数字档案馆系统中，多个平台（子系统）中的档案数据来源与其副本在数量和内容上保持一致的过程。

这里的档案数据来源是指在数字档案馆系统中完成整理、著录、封装等操作的正式档案数据，这类数据通常保存在数字档案馆局域网数字档案管理平台上，是数字档案馆系统各网络、各平台数据的来源和基准，其它平台、系统上的数据都是其副本。在 OAIS 数字档案馆系统中，这类正式档案数据称为为档案信息包（AIP），AIP 包在进入数字档案馆之前称为提交数据包（SIP），AIP 包的副本称为分发信息包（DIP）。AIP 包如果数量和内容上发生变化，对应的 DIP 也应同步这种变化，否则就可能出现不同平台间档案数量不一致、同一档案在内容不一致的情况，导致档案的不真实、不完整和不可用。

6.4 数字档案馆系统多平台数据同步的特点

6.4.1 分布式应用，一体化管理

数字档案馆的各个功能模块根据业务对象和档案数据的不同，部署在不同网络中，形成分布的应用架构。对于同

步操作同样是分布式应用，通常差异检测、同步数据的生成在局域网平台上进行，数据的分发跨越来源平台和目标平台进行，数据更新在目标平台上进行，数据更新确认则需要跨越目标平台、来源平台。这些操作在物理上隔离，且具有相对独立的控制，但总体上仍然是一个有机整体的组成部分，各平台所管理的档案数据就某一份档案而言都是一致的，即同一份档案的多个副本在不同网络 and 不同应用平台上进行处理，其管理的规则也是应是一致的，必须遵循统一的标准和规则对同步过程进行全程管理，以确保同步过程中档案数据的真实、完整、可用和安全。

6.4.2 物理隔离

数字档案馆系统在物理隔离环境中部署，使得原本逻辑上完整的数字档案馆系统，在物理上被隔离在不同的平台上，被隔离的各部分之间没有直接的通道，只能通过脱机载体、网闸等方式进行间接通讯和数据的交换。这种间接的方式，直接影响数据同步效率、同步时效和同步验证。受限于中介质的接口带宽、读写速度、存储容量以及操作速度等因素，每次同步的数据量、所需时间以及同步的频率将远低于在线方式。同时，采用间接方式进行同步结果的验证，也会影响同步完成的速度和同步结果验证的可靠性。

6.4.3 海量数据

数字档案馆系统档案数据的显著特点就大数据量。首先是数据规模大，其规模无论从存量还是增量来讲都是巨大

的。其次是单条数据量大，以图像、声音、视频等方式存储的单个文件数据量，通常是一般数据库单条记录的几百倍甚至上千倍；三是大数据的累积，与一般信息系统重点管理维护短期数据不同，数字档案馆管理的档案数据总体上是一种不断累积的数据，如同滚动的雪球一般不断增加。进行如此庞大规模数据的同步，这对同步功能的处理能力、执行效率和稳定性来说是巨大的挑战。

6.4.4 单向同步

在数字档案馆系统中，部署局域网的数字档案管理平台是数字档案管理和保存中心，它承担数字档案的管理维护和长期保存功能，其它平台的系统是它的子系统，是其功能在不同平台上的延伸。局域网保存的正式档案数据是系统中最完整，最能反映数据管理状态的数据，其它平台的数据都以它为源头，是它的副本和子集。因此，数字档案馆系统中档案数据的同步是局域网档案数据为中心，向政务网、互联网等其它平台进行单向的同步。

6.4.5 基于数据封装包的同步

在数字档案馆系统中，数据封装是维护数字档案真实、完整和可用的重要方法，数据封装包是基本的管理单元，数字档案的长期保存和分发利用都是以数据封装包为基本单位，所以数字档案馆数据同步也采用了基于数据封装包同步的方式。SIP包是来自档案形成者的原始档案数据包（包含档案实体数据和元数据），其内容通常是不允许改变的，这

是维护数字档案真实、完整和可用的基础。AIP 包用于长期保存，它包含 SIP 包的内容，同时包含数字档案馆管理过程中附加的内容，如原始档案实体的通用版本以及档号、摘要、主题、分类、开放等级、密级等管理元数据，这些管理元数据在以后还可能发生变化，如开放等级、密级等。DIP 包是面向用户和应用的数据包，也是实现数据同步需求的基础单元。DIP 包是以 AIP 包为基础根据不同的应用模板生成的副本，DIP 包可能与 AIP 完全相同，但也可能不一致。这种不一致主要是指两种包数据项目构成可能不完全相同，DIP 包的数据项目可能少于 AIP 包。但是，所包含数据的内容则必须与 AIP 包内容保持一致，这是维护数字档案真实、完整和可用的要求。

在数字档案馆系统中，档案数据包封装方式常用的主要有两类：一类的代表是《DAT/48-2009 基于 XML 的电子文件封装规范》规定的 XML 的封装，这类封装采用了签名、锁定签名等安全机制，数据的完整性、真实性、安全性保护更为严密。但是，这类封装数据处理过程相对复杂，且受限于封装对象、封装包的数据尺寸，对尺寸较大的封装对象进行封装耗时时间较长。一般适用于文书类、数据表格、静态影像等单个文件存储容量较小的原生电子档案的封装。

另一类代表是《电子档案移交与接收办法》规定的文件夹封装，这类封装一般是采用直接拷贝档案数据文件到特定文件夹的方式，不对数据进行转换操作，处理时间短，对封

装对象、封装包的存储容量基本无限制。文件夹对数据的完整性、安全性保护弱于 XML 封包方式。文件夹封装适应性较好,可对各种大小规格的数据进行封装,特别适用于音频、视频等大尺寸文件的封装。

6.5 数字档案馆多平台数据同步的目标

综合以上分析,数字档案馆多平台数据同步的目标可以界定为:在数字档案馆多平台部署的环境下,维护多平台对应档案数据的一致性,确保档案数据的真实、完整、可用和安全。

6.6 数字档案馆多平台数据同步的基本原则

6.6.1 有序可控

同步应当遵循确定的规则,按既定程序和流程进行。同步应当制定明确的规则,规则应界定同步的条件、同步的流向和同步的内容,过程应当基于规则设计,具体的操作流程应清晰、明确,可控制、可验证,同步的过程和结果应当记录并能验证和审计。

6.6.2 稳定可靠

同步功能在设计可承载范围内应能持续稳定工作,在长时间、大负荷的应用状态下同步操作应能正常运行,同步的结果应保持稳定。

6.6.3 便捷高效

档案数据基本特点就是大数据量,效率是数据同步重要要求,直接关系的到数据同步的可用性。因此,要求数据同

步能够在具体软硬件环境下达到最优的效率 ,以适应数据同步业务增长的需要。

6.6.4 安全保密

档案数据涉及国家、机构和个人的秘密、权益和隐私 ,管理过程对于信息安全有较高要求 ,这就要求实施数据同步的过程 ,必须确保数据不被损坏或篡改、信息不被泄露、不被非法复制留存。

7 数字档案馆系统档案数据分析

档案数据是数字档案馆系统的最重要的资源 ,实现档案数据的长期保存和方便利用是数字档案馆系统最主要的目标 ,档案数据管理是数字档案馆系统架构、功能设计和业务实现的出发点和重点。因此 ,要科学有序开展数据同步 ,首要的就是要识别和研究档案数据的管理和技术特点 ,以此为基础设计和实现数据同步。

7.1 云南省档案馆数字档案馆系统档案数据分析

云南省档案馆数字档案馆系统管理和存储的档案数据 ,主要来自省级立档单位移交、馆藏档案数字化和省馆采集 ,其内容主要是纸质档案数字化文件、数码照片、数字录音、数字录像及其目录 ,格式主要有 EEP、PDF、JPG、TIF、WAV、MPEG2、数据库文件等。这些数据在数字档案馆系统中 ,主要经过数据处理、入库保存、提供利用三个管理阶

段，其中入库保存和提供利用阶段已经建立相应的同步机制。

7.1.1 档案数据来源及内容分析

目前，云南省档案馆的数字档案主要来源于接收进馆的机关文书类档案数字化文件、馆藏传统载体档案经数字化加工后形成的馆藏档案数字化文件（包括扫描件、电子照片、数字化声像等）以及省馆采集的多媒体电子文件。

7.1.2 机关文书类档案数字化文件

机关文书类档案数字化文件是各立档单位将纸质档案进行全文数字化扫描所形成的 PDF 格式文件，是纸质档案的数字化副本。这些 PDF 文件按照 300dpi、24bit 真彩色扫描转化生成，通过挂接方式存入各机关的档案管理系统中。移交时，通过该系统将档案元数据及数字化档案，以电子签名封包方式，打包生成 EEP 格式的电子档案移交包，刻录在光盘内向省档案馆移交。



图 8 省级机关电子文件归档系统

7.1.2.1 馆藏档案数字化文件

馆藏档案数字化文件是省馆将馆藏档案进行数字化处理形成的文件。主要有：

文书档案扫描数字化形成的文件。一般以 TIF、PDF 格式同时保存。TIF 作为离线数据保存，PDF 格式通过挂接方式进行数字档案馆系统在线保存。

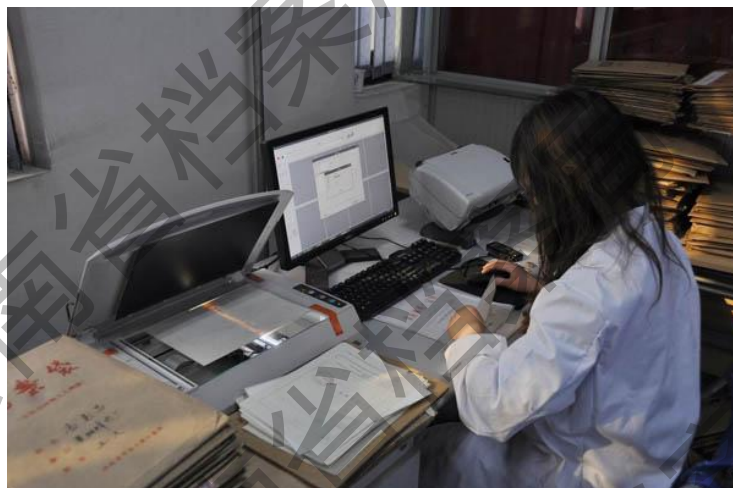


图9 馆藏档案数字化

馆藏照片档案扫描形成的文件。一般以 TIF、PDF 格式同时保存。TIF 文件作为离线数据保存，PDF 文件通过挂接方式进入数字档案馆系统在线保存。保存格式和保存方式与文书档案基本一致。

馆藏传统载体声像档案数字化文件。主要有卡式录音带数字化形成 WAV 格式音频文件，盒式录像带数字化形成 MPEG2 格式视频文件。

7.1.2.2 省馆采集的多媒体档案文件

省馆采集的多媒体档案文件 ,主要包括开展少数民族口述历史档案采集形成的 WAV 音频文件、JPG 格式数码照片、MPEG2 格式视频文件 ,收录电视新闻获得的 MPEG2 视频文件 ,进行网页归档获得的 PDF 文件。



图 10 采集声像口述历史档案

7.1.2.3 档案目录

主要是描述纸质档案、数字档案特征的目录数据 ,用于档案的标识、定位、检索、查阅等管理和利用工作。

7.1.3 档案数据管理及同步过程

档案数据进入云南省档案馆数字档案馆系统后 ,首先进行整理、编目、著录、鉴定等数据处理工作 ,然后正式入库保存 ,最后按需求分平台提供利用。在整个管理流程中 ,业务分别在不同平台完成 ,需要在不同平台间进行档案数据的同步。

7.1.3.1 数据处理

省级立档案单位移交的数字化档案、数字化部门生产的

馆藏数字化档案、采集和征集部门采集的多媒体电子档案进入数字档案馆系统后,先暂存在系统临时数据库等非正式存储区域,在这里进行相应的数据处理操作。对于机关移交数字数字化档案主要是进行整理、格式转换、编目、著录、鉴定,对于数字化部门生产馆藏数字化档案主要是进行质量验收、开放鉴定,对于采集的多媒体电子档案,主要是进行编辑、处理、编目、著录、鉴定。在处理过程中,这些数据还处于未固定的状态,是非正式的档案数据。

在云南省档案馆的数据处理工作一般在局域网平台的数字档案管理系统中进行。

7.1.3.2 入库存储

档案数据经过整理、验收、编辑、编目、著录、封装等处理,符合长期保存要求后就从临时库进入正式库进行保存。在这里数据的内容和形式已经相对固定,成为正式数据,可以正式提供利用。

云南省档案馆的正式档案数据在一般在局域网平台、屏蔽网平台的数字档案管理系统进行存储和管理。

7.1.3.3 提供利用

以正式数据为来源,根据利用对象、利用目的不同生成对应的利用副本,分别同步到局域网、政务网、互联网平台上提供用户利用。

云南省档案馆的利用副本数据一般在局域网、政务网、互联网的档案利用系统中进行储存。

7.1.3.4 数据同步

目前，云南省档案馆的数据同步流程主要有两类：

一是局域网数字档案管理平台正式库数据向屏蔽网存储平台备份库的完整同步。

同步通过独立的同步工具软件进行，以全宗为单位将目录数据和对应的原文数据从局域网数字档案管理平台的正式库完整拷贝到移动硬盘，然后将移动硬盘接入屏蔽网核心存储平台，使用同步软件将档案数据导入备份库完成同步。

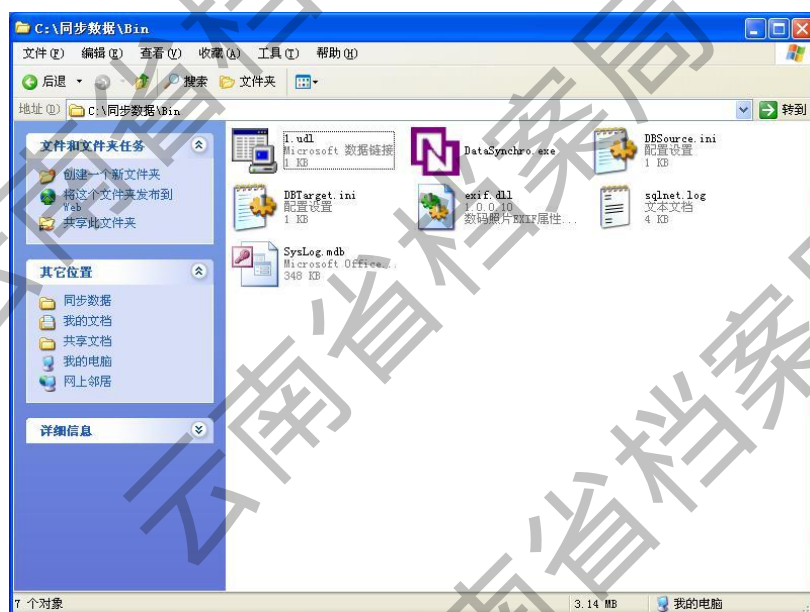


图 11 屏蔽网数据同步软件

二是数字档案管理平台正式库档案数据向政务网、互联网利用平台利用同步。

同步通过数字档案管理平台的数据分发功能进行，以文件为单位将文件目录和对应原文数据封装成 EEP 格式的数据包，用移动介质将 DIP 包拷入相应利用平台，通过利用平台的数据接收功能完成同步。



图 12 DIP 数据包导出

现有的同步机制主要存在以下问题：

一是缺乏直观和清晰的管理和控制。同步操作没有授权和审核机制，同步过程、数量、流向没有直观、清晰的反馈，同步日志无法直接查，管理人员无法对同步进行有效监督，操作人员难于对过程进行控制。

二是缺乏及时方便维护来源数据和副本数据一致性的控制机制。对来源数据变化只是被动的记录，不能自动提示，来源数据的变化难以及时同步到副本数据，从而导致两者内容上的不一致。

三是同步方式不灵活。只能以全宗或者文件为单位进行同步，不能灵活选择同步单位、同步数据范围，对局部、批量数据的变化无法方便选择和及时同步；完整同步只能使用移动硬盘，同步控制数据存储在移动介质中存在一定安全风险。

四是同步效率较低。受限于半人工操作方式和脱机介质读写速度，同步操作较为繁琐，数据读写速度缓慢，同步效率较低。

7.2 数字档案馆系统档案数据类型及特点

数字档案馆系统的档案数据的可以从数据构成、功能用途、逻辑结构等角度进行区分。从数据构成可以区分为数字档案文件数据和数字档案元数据；从功能用途可以区分为提交数据、存档数据和分发数据；从逻辑结构可以区分为结构化数据和非结构化数据。

7.2.1 数据构成

7.2.1.1 数字档案文件数据

数字档案文件是指以计算机文件保存的数字档案，是档案数据的主要内容。数字档案文件从其信息记录和表现的形式来区分，可以区分为文本、数据表格、图形，静态影像、音频、视频等类别。

文本类文件主要包括 TXT 纯文本、文字处理软件产生的文本如 WORD 等流式文件以及 PDF 版式文件。WORD 文件主要来源于立档单位的原始电子文件，一般要求转换为 PDF 格式随原始电子文件一同存档。PDF 版式文件由立档单位 OA 系统中直接形成、由其它原始格式（如网页、公务电子邮件、WORD 文档、电子表格、演示文稿等）转换形成或由传统纸质档案经数字化加工后形成。

数据表格类文件产生于立档单位的业务系统或办公软

件，属于流式文件，如 EXCEL 电子表格。

WORD、EXCEL、TXT 等流式文件以字符为处理单位，文件格式简单，体积较小，比较易于进行数据同步。

静态影像类文件包括征集进馆或由档案馆采集、拍摄的电子照片，传统纸质、胶片档案经数字化加工后形成的数字副本，通用格式有 JPEG、PDF 和 TIFF。这类文件结构相对复杂，尺寸较大，且大量数字档案以这种格式保存，在进行大数据量同步时，对同步的效率、稳定性要求较高。

图形类文件主要来源于基建档案、工程图纸等科技档案中的设计文件，一般产生于立档单位的业务系统，如 CAD、VISIO 等。

音频类文件包括征集进馆及由档案馆采集、录制的原生的录音类电子档案，录音带等传统载体录音档案经数字化加工生成的数字副本，通用格式是 WAV、MP3。

视频类文件包括征集进馆的录像类原生电子档案，由档案馆自行摄制或从 DVD、VCD 播放光盘、电视节目采集形成的录像类电子档案以及录像带等传统载体录像档案经数字化加工的形成的数字副本。通用格式是 AVI、MPG。

文本、静态影像类文件单个文件占用存储空间相对较小，可采用 XML、文件夹等方式封装，技术适应性较好。图形、音视频类文件单个文件占用存储空间较大，传输速度慢，不宜采用 XML 封转，一般使用文件夹方式进行封装。

7.2.1.2 数字档案元数据

元数据是指“描述文件的背景、内容、结构及其整个管理过程的数据”。其主要功能是：捕获与保存文件背景信息；系统与文件结构的保存；相关描述信息的生成与保存；鉴定与处置数据集成；文件生命周期的管理；系统功能的保存与迁移；组织机构信息资源的检索与定位系统的创建。数字档案元数据解释数字档案文件的技术和内容结构，记录数字档案形成和管理过程，揭示数字档案特征，是管理、定位和检索数字档案文件，维护电子档案真实性、完整性、可用性的基础。

我国档案行业标准《DA/T 46-2009 文书类电子文件元数据方案》把电子文件元数据从概念层次上区分为文件实体元数据、机构人员实体元数据、业务实体元数据和实体关系元数据四个域。

文件实体元数据主要记录文件、档案实体特征的描述信息。包括聚合层次、来源、档号、内容描述、形式特征、电子属性、数字化属性、电子签名、存储位置、权限管理、附注等 75 项。

机构人员实体元数据主要记录对文件、档案进行操作的相关机构、人员描述信息，包括机构人员类型、机构人员名称、组织机构代码、个人职位四项。

业务实体元数据主要记录对文件、档案进行操作的具体行为的描述信息，包括业务状态、业务行为、行为时间、行为依据、行为描述等五项。

实体关系元数据主要记录文件、业务、机构人员等实体的内、外部关系的描述信息，包括实体标识符、关系类型、关系、关系描述四项。

7.2.2 功能用途

7.2.2.1 提交数据

提交数据是档案数据生产者提交给数字档案馆系统，等待进行整理、验收、编辑、编目、著录等处理的档案数据。提交数据的来源包括立档案单位移交的电子档案和数字化档案、档案馆数字化部门生产的数字化档案、档案馆采集和征集部门采集的各种形式电子档案以及因管理需要重新进行错误修正、格式转换、鉴定、整理等操作的存档数据。提交数据是未固定的非正式数据，一般不提供利用，通常在数字档案馆的数据处理系统上存储。对应 OAIS 模型，提交数据的概念和功能与 SIP 包类似。

7.2.2.2 存档数据

存档数据是指完成整理、验收、编辑、编目、著录、封装等操作，符合长期保存要求，进入正式数据库进行保存的正式档案数据。存档数据在保存期间，可能因为管理原因对其部分元数据进行修改，如保馆期限、密级、开放等级等，这时存档数据重新转换为提交数据，待处理完成后再次转换为存档数据存档。存档数据相对固定且具备档案功能的数据，是数字档案管理、利用的核心，是数据同步的起点和基准。存档数据通常在数字档案馆的数据长期保存和备份系统

上进行存储。

对应 OAIS 模型 ,存档数据的概念和功能与 AIP 包类似。

7.2.2.3 副本数据

副本数据是指以正式数据为来源 ,根据不同用途生成的正式数据副本 ,如用于提供查阅的利用数据及用于异地备份的备份数据。副本数据可能包含全部或者部分的存档数据 ,如用一些利用者不需要的数据及馆内管理过程形成的部分元数据就不应复制到利用数据中。对应 OAIS 模型 ,副本数据的概念和功能与 DIP 包类似。

7.2.3 逻辑结构

7.2.3.1 结构化数据

结构化数据即行数据 ,就是存储在数据库里 ,可以用二维表结构来逻辑表达实现的数据。简单的说 ,所有关系型数据库中的数据全部为结构化数据。在数字档案馆系统中 ,档案数据的元数据一般以两种方式保存 :一种是以非结构化数据文件方式保存在文件系统中 ,如 XML 文件 ;一种是以结构化数据形式存储在数据库中。非结构化的文件存储主要是为降低数据对系统的以来 ,确保的元数据的的长期可用 ,结构化的数据库存储则更利于数据操作的高效、稳定和可控。档案元数据的管理和维护通常在数据库系统中进行 ,元数据的变化首先在数据库系统中通过增、删、改操作完成 ,然后写入 XML 文件形成新的数据包 ,替换原始数据包进行存储 ,从而实现元数据的便捷管理和长期保存。

7.2.3.2 非结构化数据

相对于结构化数据,不方便用数据库二维逻辑表来表现的数据即为非结构化数据,包括所有格式的办公文档、文本、图片、XML、HTML、各类报表、图像和音频/视频信息等。数字档案文件、档案信息包都是典型的非结构化数据。这类数据结构多样、展现形式直观、数据量巨大且增长较快,是数字档案馆存量最大的数据。在数字档案馆系统中,数字档案提交、存储和分发利用都是以信息包这种非结构化数据为单元实施的。因此,数据同步也是以非结构化数据文件的形式进行。

8 档案数据同步的管理要求

数字档案馆多平台部署环境下,由于管理、利用和安全等业务的需要,必须进行同步以维护多平台档案数据的一致性,同时同步过程必须遵行并符合有关法律规范的要求,以确保档案数据的真实性、完整性和有效性。

8.1 业务要求

在数字档案馆系统,对数字档案的管理和利用是一个动态的过程,在此过程中档案数据的稳定是相对的,而变化则是必然的,这些变化包括对数据的增加、删除和对数据内容的修改等。因此,必须通过同步将这些变化反映到数据的所有副本上,以维护分布存储档案数据的一致性,以确保档案

管理和利用的一致、有效和可用。同时，为了确保安全，必须采取措施来确保同步过程和同步数据的安全。

8.1.1 数据增加

数据增加主要是通过接收、挂接、采集等方式增加档案数据，数据增加通常导致数据数量的变化。

8.1.2 数据删除

数据删除主要是删除系统中已有的档案数据，数据删除通常导致数据数量的变化。

8.1.3 数据修改

档案数据进入数字管理平台后，为有效、科学进行管理，需根据《电子档案移交与接收办法》、《全宗规范》、《档案编制规范》、《档案著录规则》、《归档文件整理规则》、《电子档案归档与管理规范》等相关要求，对数据进行检验、编号，法律效力鉴定（真实性、完整性、有效性鉴定），密级鉴定、保管期限鉴定、开放鉴定等操作；同时，为了确保数字档案的长期可读，可能需要对一些接收非通用格式的电子档案进行格式转换，这些操作将使数字档案元数据和文件数据发生变化。

8.1.3.1 数据检验

对进馆（或馆藏）数据的检验，检查元数据信息与电子文件信息的一致性，纠正元数据与电子文件对应错误；纠正元数据信息错误，特别是题名、年度、机构、密级等关键信息与电子文件不符；纠正电子文件损坏、分辨率或格式不符

合要求等造成文件不能正常读取的错误。由于电子档案数量巨大，错误不易发现等特点，数据检验过程应贯穿于电子档案收集、保存、利用的整个过程，在任何业务过程发现错误，应通过合法有效的管理机制予以纠正。

8.1.3.2 档号调整

对移交的电子档案按照综合档案馆管理要求进行重新编号。部分项目档案、专业档案、专题档案、新档案类型等，需要对其重新编号，以便保管、识别和查找。调整编号，需直接对元数据中编号信息进行调整和补充。

8.1.3.3 法律效力鉴定

对电子档案反映真实性、完整性、有效性信息的元数据和电子文件信息进行鉴定，发现信息缺失、不规范等情况的，应及时根据相关规定对元数据信息加以补充，确保其法律效力。

8.1.3.4 密级鉴定

对密级确定不准或对保密有新要求、新规定时，需对原鉴定的密级进行补充，使其满足保密相关要求。

8.1.3.5 保管期限鉴定

对未按要求划分保管期限的，应按照规定要求进行重新划分，并将相关信息在元数据中标识。

8.1.3.6 开放鉴定

对电子文件的开放级别、公开范围等情况进行鉴定，以确定公开范围，利用程度等，需对元数据信息进行标识。

8.1.3.7 格式转换

对数字档案文件格式进行检测,将参数不规范或者格式通用性较低的文件按统一标准转为统一的格式,与原始格式数据一同保存,以确保数字档案原始性的同时提高其可用性。

通过以上在档案管理业务中能引起数据变化因素的分析,能有效把握在管理业务过程中数据变化的因素,并将这些变化原因、过程和结果在元数据中加以保存,确保不因业务管理过程元数据记录的缺失,导致数据失去合法性。通过管理业务产生的元数据的完善,保证电子文件的原始性、真实性、有效性,为数据同步到各业务平台提供利用奠定基础。

8.1.4 利用

档案的保管是为了有效利用,在数字档案馆系统中应根据不同服务对象和不同档案的开放范围,在相应的服务平台提供相应等级的数字档案信息资源利用共享服务。具体来说局域网平台面向档案馆工作人员和来馆利用人员,政务网平台面向本级党政机关人员,互联网平台面向互联网公众。应按照《保密法》、开放鉴定管理规定等,对数据进行科学的划分,从密级信息、年代信息、档案具体内容是否涉及敏感信息等几个方面进行判断,科学划分同步数据的流向、范围和内容,形成针对不同业务平台数据同步策略和同步数据模板。

8.1.5 安全管理

安全是档案管理的基本要求,安全保障体系建设是数字档案馆建设的基础性工作。数字档案馆的安全包括数字档案数据的安全和信息系统及其网络平台的安全。数据安全就是要保证数字档案信息的可靠、可用、不泄密、不被非法更改等。系统及其网络平台安全就是要保持系统软硬件的稳定性、可靠性、可控性。

数字档案馆档案数据同步的安全要求主要体现在四个方面:一是系统安全,用于同步的软、硬件系统的安全等级应与所同步数据的安全要求相适应;二是规则安全,同步应当基于明确的安全原则,即高安全等级的数据不应同步到低保护等级的平台上;三是过程安全,应建立安全审核控制机制,同步操作应当遵循明确的规则并经过授权和审批后进行,同步过程和结果应有监管和审计;四是技术安全,应采用身份识别、加密、防篡改、介质管控等相应技术措施,以确保同步过程和数据的安全。

8.2 法规要求

8.2.1 《中华人民共和国电子签名法》

《中华人民共和国电子签名法》(以下简称《电子签名法》)由中华人民共和国第十届全国人民代表大会常务委员会第十一次会议于2004年8月28日公布,自2005年4月1日起施行。

《电子签名法》首次明确了电子文件具有法律效力,电子签名与传统手写签名和盖章具有同等的法律效力,这在一

定程度上解决数字档案的法律地位和作用问题。同时,《电子签名法》对电子文件具备法律效力的条件也做规定,集中体现在电子文件的原始性、真实性和有效性等方面,这是确定数字档案和数字档案管理活动合法性的重要依据:

8.2.1.1 原始性要求

《电子签名法》第五条规定:“符合下列条件的数据电文,视为满足法律、法规规定的原件形式要求:(一)能够有效地表现所载内容并可供随时调取查用;(二)能够可靠地保证自最终形成时起,内容保持完整、未被更改。但是,在数据电文上增加背书以及数据交换、储存和显示过程中发生的形式变化不影响数据电文的完整性”。

8.2.1.2 真实性要求

《电子签名法》第八条规定:“审查数据电文作为证据的真实性,应当考虑以下因素:(一)生成、储存或者传递数据电文方法的可靠性;(二)保持内容完整性方法的可靠性;(三)用以鉴别发件人方法的可靠性;(四)其他相关因素”。

8.2.1.3 有效性要求

电子文件的有效性标准主要是从保存、移交和接收两个环节加以保障的。

保存环节:《电子签名法》第六条规定:“符合下列条件的数据电文,视为满足法律、法规规定的文件保存要求(一)能够有效地表现所载内容并可供随时调取查用;(二)数据

电文的格式与其生成、发送或者接收时的格式相同，或者格式不相同但是能够准确表现原来生成、发送或者接收的内容；(三)能够识别数据电文的发件人、收件人以及发送、接收的时间”。

移交和接收标准环节：《电子签名法》第九条规定：“数据电文有下列情形之一的，视为发件人发送：(一)经发件人授权发送的；(二)发件人的信息系统自动发送的；(三)收件人按照发件人认可的方法对数据电文进行验证后结果相符的”。《电子签名法》第十条规定：“法律、行政法规规定或者当事人约定数据电文需要确认收讫的，应当确认收讫。发件人收到收件人的收讫确认时，数据电文视为已经收到”。

8.2.2 《电子文件管理暂行办法》

《电子文件管理暂行办法》(中办国办厅字〔2009〕39号)(以下简称《办法》)于2009年9月28日由中共中央办公厅、国务院办公厅联合下发。《办法》是我国第一部电子文件管理的中央层面的行政规章。《办法》确立了电子文件全程管理的原则，明确电子文件管理机构和职责，明确规定“各级国家综合档案馆负责接收和保管本馆接收范围内各单位形成的具有永久保存价值的电子文件，并依法提供利用”，对电子文件管理的形成办理、归档移交、保管利用全过程提出了明确的管理要求，是目前我国电子文件、电子档案管理的基本规范。

《办法》要求“电子文件及其元数据自形成起完整无缺、来源可靠，未被非法更改”，“电子文件传输、交换时，应当遵循相关要求，对传输、交换过程予以记录”，“电子文件及其元数据应当同时归档”，“属于国家综合档案馆接收范围的电子文件，应当按照规定时限向同级国家综合档案馆移交”，“反映电子文件保管、利用过程的相关信息应当记录和保存”，“对有价值的电子文件提供分层次、分类别共享应用”，“应当为利用者提供真实、可靠的电子文件，并采取有效措施确保电子文件不受损害”。

8.2.3 《电子文件归档与管理规范》

《电子文件归档与管理规范》(GB/T18894-2002)(以下简称《规范》)是我国电子文件归档管理的国家标准。

《规范》提出电子文件管理“真实性、完整性和有效性”的目标和原则，明确“应对电子文件的形成、收集、积累、鉴定、归档等实行全过程管理与监控，保证管理工作的连续性”，“保证电子文件的凭证作用”。《规范》规定在公务活动中产生的，具有保存价值的电子文件的形成、积累、归档、保管、利用、统计的一般方法，是电子文件管理的基本指南。

《规范》对电子档案管理各业务环节的要求更具体，针对性、可操作性更强，与《电子签名法》中有真实性、完整性和有效性保障要求总体上是一致的，将他们有机结合在一起，将在多平台数据同步过程中确保电子档案数据能发挥法律效力起到促进作用。

8.2.4 《电子档案移交与接收办法》

《电子档案移交与接收办法》由国家档案局（档发[2012]7号）于2012年8月29日发布实施，《办法》对移交、接收、保管电子档案提出了具体的要求。

在移交工作中，要求元数据应当与电子档案一起移交，应采用基于XML的封装方式组织档案数据；电子档案有相应的纸质、缩微制品等载体的，应当在元数据中著录相关信息；所移交的数据为非加密、非压缩格式。从移交方式上，可采用离线和在线两种方式，其中，离线移交电子档案移交单位一般采用光盘移交电子档案，光盘应当符合归档要求；在线移交电子档案的，传输的数据应当包含符合要求的电子档案及其元数据。

在接收管理中，要求建立电子档案接收平台，进行电子档案数据的接收、检验、迁移、转换、存储等工作。要求对所接收电子档案数据的准确性、完整性、可用性和安全性进行检验。要求将电子档案交接、迁移、转换、存储等信息补充到电子档案元数据中。对电子档案数据迁移和转换操作时，要求对操作前后的数据一致性进行校验。

《办法》中的要求使电子档案在移交、接收工作中如何确保能证明原始性、真实性、有效性的相关信息有效收集，提出了根据业务需求对元数据信息的补充完善的要求，提出了在电子档案和元数据信息进行迁移时，必须对数据进行一致性校验的要求，为跨平台数据同步业务实施中提供了较为

明确的业务流程依据。

9 数据同步的技术要求

数据同步最终是依靠数字档案馆计算机系统的相应的数据同步处理来实现和完成的，其要求体现在功能、性能和安全的三个方面。

9.1 功能要求

9.1.1 支持集中式管控

集中式管控就是按照数字档案馆系统一体化管理的原则，设置专门的同步管理和控制系统和同步数据库，以存档数据的管理平台为中心，协调和控制分发数据生成以及其它平台的同步操作，并记录和验证同步操作的过程和结果。

9.1.2 支持自动或者手动同步

自动同步就是系统根据同步规则，自动选择相应的符合同步条件的档案数据，按照数据模板生成同步数据，向目标平台实施同步。而手动操作就是由人工方式选定同步数据，确定同步流向和同步数据模板，启动同步操作的方式。

9.1.3 支持离线或在线同步

离线同步是指在物理隔离的平台之间，使用中间介质，以半人工方式实现的数据同步。而在线同步是指在逻辑隔离或者建立了数据通道的平台之间，通过数据链路直接进行的数据同步。为了适应不同的网络隔离情况，数据同步应当同

时支持离线和在线方式的同步。

9.1.4 支持多种封装方式

支持选择 XML、文件夹等不同方式进行数据封装迁移。用户能够根据数据格式、存储容量等特性，选择不同方式进行数据封装同步。

9.1.5 支持多种脱机介质和网闸设备

数据同步操作能够使用光盘、硬盘、U 盘常见的脱机介质和网闸设备，通过不同硬件接口以不同数据传输速度完成同步操作，同步过程中系统能对这些脱机介质直接进行读写操作，无需而人工辅助进行数据的拷贝复制。

9.2 性能要求

同步操作的性能要求主要体现在同步速度、可控性、可靠性、稳定性、安全性等方面。

9.2.1 同步速度

同步速度，就是在给定系统软件、硬件环境下，单位时间内能够完成同步操作的数据量。在多平台环境下的同步操作需要经过分发数据生成、写入同步介质、同步介质运送、读取同步介质、数据更新、同步确认等多个环节，同步速度受到同步数据的数量和单条数据的容量、分发数据的生成速度、同步介质的读写速度、同步介质送达速度、数据更新速度、同步确认速度等多重因素的影响。因此，同步速度是一个相对的综合测算值，必须在特定的环境下，综合测算各环节的速度才能最终得出。

9.2.2 可靠性要求

可靠性是指数据同步功能在某个时间点能够正常运行，运行的过程和结果达到预期目标。可靠性主要体现同步结果正确率，一般要求平均运行正确率大于 99%，对于单条数据而言必须达到 100%。同时，对于出错或者没有完成的同步，能够检测并能自动重新同步，直至同步正确或者取消同步。

9.2.3 稳定性要求

稳定性是指数据同步功能在一定时间内正常运行。由于，数据同步是数字档案馆系统正常运行的基础业务，因此稳定性的要求应与整个数字档案馆系统对稳定性的要求一致。

9.2.4 可控性

可控性是指数据同步过程和结果可控制、可追溯、可是审计。数据同步建立操作权限控制机制，操作过程和操作结果的记录机制，以及记录保存和审查机制。

9.2.5 安全性

安全性主要是指数据同步过程能确保数据的安全和完整。安全主要是指档案数据在同步过程不被泄露、不被篡改。完整是指档案数据在同步过程组成要素完整无缺失。

10 数字档案馆多平台档案数据同步管理的业务模型

综合以上研究,本课题提出了数字档案馆多平台档案数据同步管理的基本业务模型(以下简称同步模型),描述了数据同步涉及的功能、数据以及过程。

10.1 模型描述

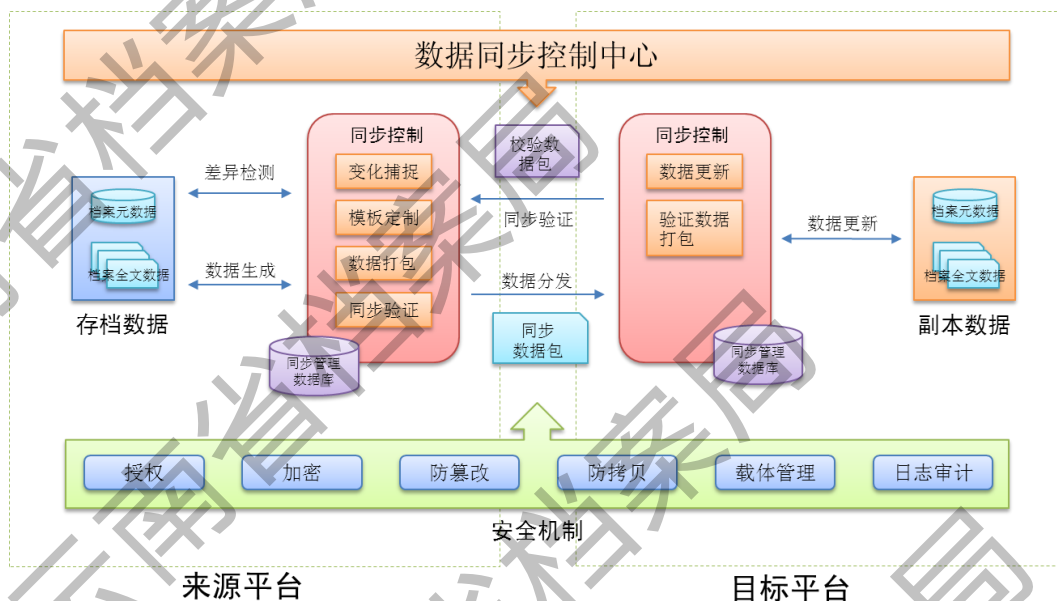


图 13 数字档案馆多平台档案数据同步业务模型

同步模型由分布在来源平台和目标平台的同步控制中心,模板定制、变化捕捉、数据封装、数据更新、校验数据封装、同步验证等六个同步功能模块,同步管理数据库,存档数据、副本数据两类数据,同步数据包、验证数据包两类中间数据包、安全机制构成以及同步流程构成。同步流程分为首次同步和变化同步两个阶段,同步流程主要包括变化捕捉、同步数据生成、数据分发、数据更新、确认数据生成、同步验证等步骤。

10.1.1 数据同步控制中心

数据同步控制中心是数据同步的管理和控制平台,集成数据同步各环节操作功能和菜单,集中显示操作过程反馈的各类信息,是数据同步过程操作的入口和信息反馈的出口。同步控制中心采用分布式部署在来源平台和目标平台。

10.1.2 同步控制功能模块

10.1.2.1 模板定制

提供用户根据不同用途预先设定同步数据包内容和机构的功能。模板规定不同用途(流向)的同步数据包的内容和结构,为同步数据包的生成提供标准。同步数据包应包含同步管理数据(包括档案唯一性标识、同步流向、数据变化标记、档案数据完整性校验值等)、档案数据(包括数字档案原文数据、档案元数据)。档案元数据的选择根据不同平台数据应用的需求确定。

10.1.2.2 变化捕捉

提供对来源数据的增、删、改等变化进行检测并记录的功能。检测的内容包括了数据的增加和删除,数据内容(元数据、原文数据)的增加、删除、修改。捕捉和记录的粒度最大不能超过文件级。变化捕捉的实现方式应综合考虑粒度、数据特征、系统资源开销因素等。

10.1.2.3 数据封装

提供同步数据选择、设定模板和封装功能。

同步数据选择分为两种方式:一种是人工选择需要同步数据,主要用于同步新形成的档案数据;一种是自动根据设

定的规则自动筛选需要同步的数据 ,主要用于同步后又发生变化的数据以及符合同步规则的数据。

应用模板就是为已选出的数据指定同步数据包模板。

封装则是根据指定的模板生成同步数据包的过程。封装生成的数据包在物理隔离环境需要复制到脱机载体上进行数据分发。

10.1.2.4 数据更新

提供在目标平台上同步数据包提取和更新副本数据的功能。

10.1.2.5 校验数据封装

提供生成校验数据和封装的功能 ,为来源平台确定同步完成提供反馈。以更新的数据为基础 ,生成校验数据并封装。

10.1.2.6 同步验证

提供确认同步结果的功能。在来源平台上通过比对校验数据 ,以判断同步是否正确完成。

10.1.3 同步管理数据库

用于记录同步过程形成各种元数据。同步管理数据库以文件为单位记录每一份档案的唯一性标识、已同步流向、同步完成标记、数据变化标记 (增、删、改)、档案数据完整性校验值等。

10.1.4 来源平台和目标平台

10.1.4.1 来源平台

来源平台通常是指数字档案馆多个平台中处于核心位

置的平台,通常是具备完整数字管理功能的数字档案管理平台(系统),一般在数字档案馆的局域网上部署,与其它系统和网络物理隔离。

10.1.4.2 目标平台

目标平台通常是指数字档案馆中或部分承载业务功能的平台,如政务内网、政务网外网、互联网的利用平台、屏蔽网的备份平台以及远程异地备份平台等。

10.1.5 存档数据和副本数据

10.1.5.1 存档数据

存档数据是存储在数字档案馆来源平台中长期保存的正式的档案数据。它的内容最完整的,状态最新。存档数据是同步的对象。

10.1.5.2 副本数据

副本数据存储在目标平台,是以存档数据为蓝本按照不同用途复制生成的存档数据的副本。副本数据是存档数据的子集,根据用途不同在内容一般与存档数据相同或者少于存档数据。在完成同步更新之前,副本数据的更新状态一般晚于存档数据。副本数据体现同步的结果。

10.1.6 中间数据包

中间数据包,是同步过程中用于档案数据同步分发和确认信息交换的数据。

10.1.6.1 同步数据包

同步数据包是以存档数据为蓝本根据同步模板生成的

用于同步分发的数据包。一般包含了同步管理数据(包括档案唯一性标识、同步流向、数据变化标记、档案数据完整性校验值等)、档案数据(包括数字档案原文数据、档案元数据)。用于备份用途的同步数据包的内容构成和数据结构与存档数据一致。用于利用等其它用途的其内容构成和数据结构与存档数据可能会有所不同。

10.1.6.2 校验数据包

校验数据包用于确认同步更新完成且结果正确的信息反馈数据包。校验数据包在目标平台生成,包含了档案唯一性标识、同步流向、数据变化标记、档案数据完整性校验值等数据。同步系统通过比较校验数据包和同步管理数据库的数据来确认同步是否正确完成。

10.1.7 安全机制

安全机制通过管理和技术的方法,确保同步操作的过程及数据安全。

10.1.7.1 权限控制

对同步操作的权限进行管理和控制,同步操作应经过必要的授权和审批程序,包括功能授权、操作授权等。

10.1.7.2 加密

对生成的同步数据包进行加密,防止其中信息泄露。

10.1.7.3 防篡改

采用封包、加密和完整性验证等方法,防止同步数据的内容被篡改。

10.1.7.4 防拷贝

采用防拷备载体、数据定时自毁等技术防止同步数据包被非授权的拷贝和留存。

10.1.7.5 载体管理

对用于同步数据交换的载体进行识别和管理,并明确载体使用的规则,防止由于载体管理失控对数据和系统带来的安全风险。

10.1.7.6 日志审计

同步操作的全过程都应进行日志记录,并提供日志查阅和备份功能,以进行同步操作的审计。

10.1.7.7 同步过程

同步过程分为首次同步、变化同步两个阶段。首次同步是指档案数据首次从来源平台向目标平台的同步。而变化同步是指已经进行过同步的数据,由于自身发生变化而再次进行的同步。

10.1.7.7.1 首次同步

10.1.7.7.1.1 定制模板

根据同步数据用途,预先设定不同用途的同步数据包格式。

10.1.7.7.1.2 选择数据

通过两种方式进行数据选择,一种手动选择数据,通过点选、划定范围等方式选择一条或多条数据。另外一种是通过对设定规则自动方式筛选出需要同步的数据。数据选定后,

确定其同步数据包的模板和同步流向。

10.1.7.7.1.3 生成同步数据

根据不同的分发模板生成对应的同步数据包,将同步数据数据包管理数据写入同步管理数据库,包括档案标识、同步流向、完整性校验值等。

10.1.7.7.1.4 分发同步数据

将同步数据包写入脱机载体,移送目标平台。

10.1.7.7.1.5 更新数据

在目标平台上,读取同步数据包中的同步管理数据和档案数据,并与本平台同步管理数据进行比对,完成档案数据的增、删、改操作,将本次数据更新情况记入本平台同步管理数据库。

10.1.7.7.1.6 生成校验数据

将本次数据更新情况封包形成校验数据包(包括档案唯一性标识、同步流向、数据变化标记、档案数据完整性校验值等)。

10.1.7.7.1.7 返回同步数据

将同步数据包写入脱机载体,返送回来源平台。

10.1.7.7.1.8 同步验证

在来源平台,读取校验数据并与同步管理数据比对,如果比对一致则,写入同步成功标识,结束本次同步。如果比对不成功,则写入同步失败标识,重新启动同步流程。

10.1.7.8 变化同步

10.1.7.8.1 变化捕捉

对来源平台档案数据的增加以及已同步数据的删、改等变化进行检测并形成数据更改记录，存入同步管理数据库。

10.1.7.8.2 选择数据

比对数据更改记录和数据同步记录，自动筛选出同步后又发生变化的数据，自动设定这些数据的同步数据包模板和同步流向。

后续过程与首次同步一致，参见 10.1.7.7.1.3—10.1.7.7.1.8 内容。

10.2 模型验证

基于上述模型，本课题研究开发了验证软件，对同步模型的主要内容进行验证，确认了模型的可行性和可用性。

10.2.1 软件简介

软件通过来源平台和目标平台两个控制界面，进行数据同步操作。

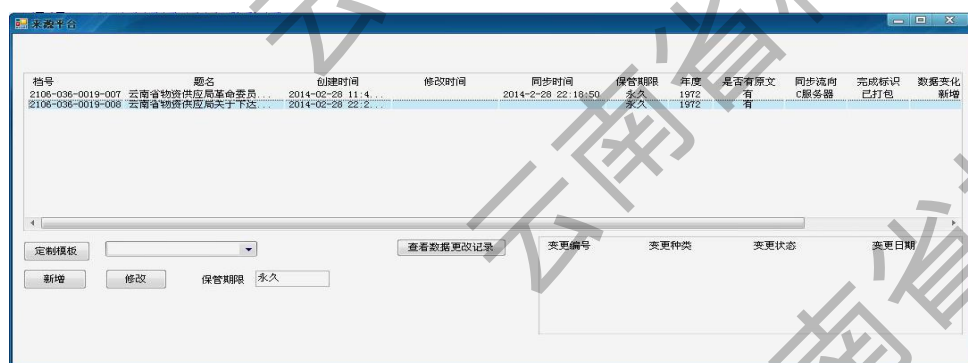


图 14 来源平台界面



图 15 目标平台界面

10.2.1.1 硬件环境

当前主流配置普通 PC 机。

10.2.1.2 软件环境

操作系统：Microsoft XP、Microsoft Vista、Microsoft Windows2003。

编译环境：Microsoft visual C++ 6.0 编译，使用时需要安装 Microsoft .Net framework 3.5 以上版本。

数据库：sqlite 数据库，已集成到应用程序，使用时不需再次安装。

10.2.1.3 执行程序目录简介

来源平台：

来源平台执行程序位置：

来源平台\bin\Release\WindowsFormsApplication1.exe

来源平台生成数据包存放目录：

来源平台\bin\Release\send\

来源平台接收目标平台确认数据包存放目录：

来源平台\bin\Release\recv\

目标平台:

目标平台执行程序位置 :

目标平台\bin\Release\WindowsFormsApplication1.exe

目标平台确认数据包存放目录 :

目标平台\bin\Release\send\

目标平台接收来源平台生成数据包存放目录 :

目标平台\bin\Release\recv\

10.2.2 流程简介

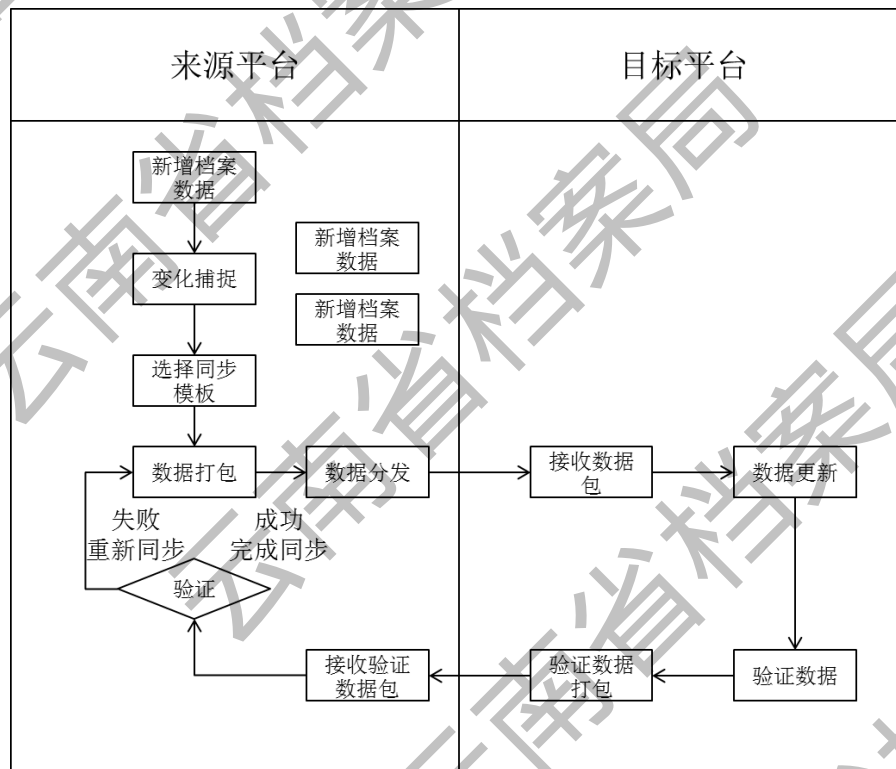


图 16 系统程序流程图

10.2.2.1 模板选择

业务说明 :为系统中存储的档案条目指定一个已定制好的模板文件，通过该模板文件，设定分发数据的结构、流向等相关信息。

操作说明 :在“来源平台”界面点击模板文件下拉列表选

择模板文件，在档案条目列表选择档案条目数据，打开右键菜单，点击“选择模板”，完成模板文件选择。

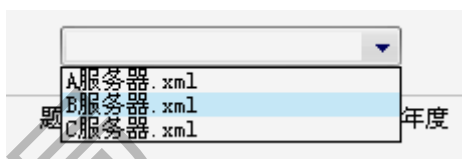


图 17 模板文件下拉列表

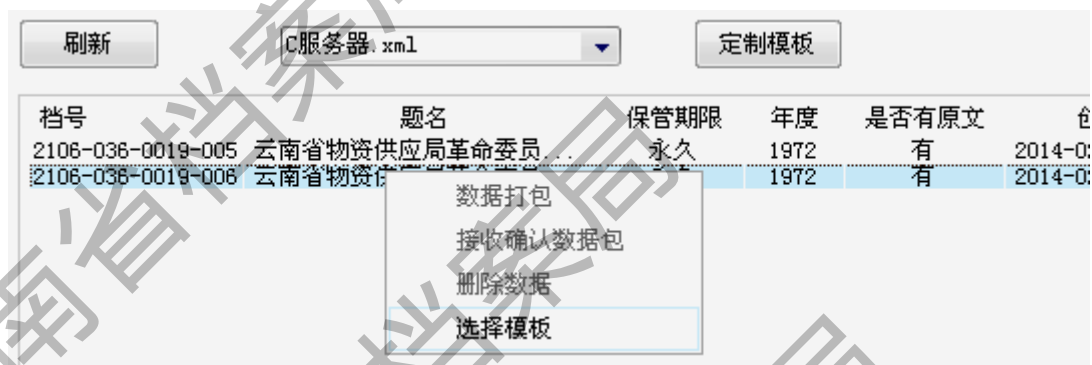


图 18 选择档案条目右键菜单

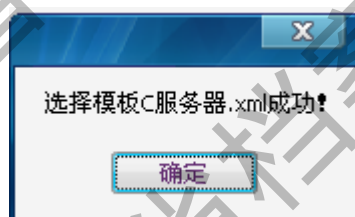


图 19 选择模板成功后系统提示信息

10.2.2.2 数据打包

业务说明：数据打包功能根据不同的分发模板生成对应的同步数据包，并将同步数据包写入，移送目标平台。

操作说明：在“来源平台”界面档案条目列表选择档案条目数据，打开右键菜单，点击“数据打包”，完成所选档案条目的数据打包，打包完成后，档案条目列表中显示同步时间、同步流向、完成标识、数据变化标识、完整性校验等信息。

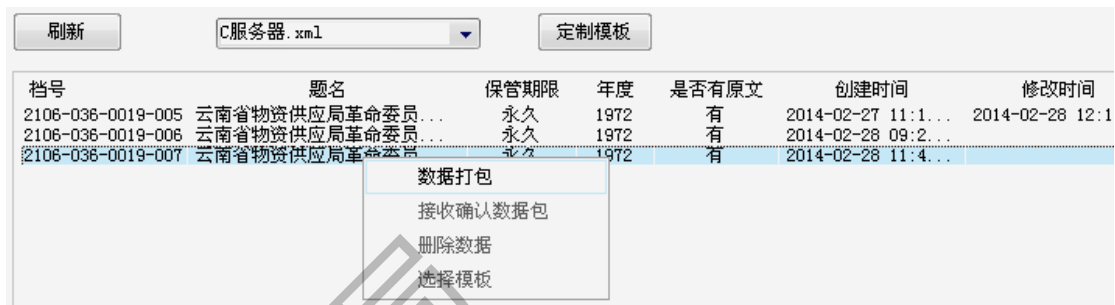


图 20 选择档案条目右键菜单



图 21 数据打包成功后系统提示信息

同步时间	同步流向	完成标识	数据变化标识	完整性校验
2014-2-28 9:27:17	A服务器	已打包	新增	857579

图 22 数据打包成功后列表显示信息

10.2.2.3 档案数据包移动

业务说明 :使用人工方式通过脱机载体将生成在来源平台的档案数据包移动到目标平台。

操作说明 : 拷贝来源平台“\bin\Release\send\”目录下的档案数据包“档号.zip”.移动到目标平台“\bin\Release\recv\”目录下。

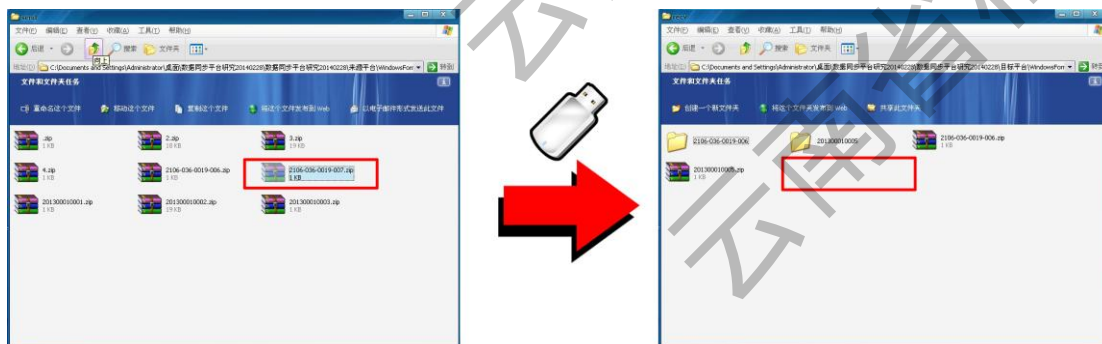


图 23 移动档案数据包示意图

10.2.2.4 数据更新

业务说明 :数据更新功能读取通过脱机载体移送到本目标平台分发数据包中的同步管理数据 ,将此读取到的同步数据与本目标平台数据库中的同步管理数据进行比对 ,并根据比对结果进行更新操作。

操作说明 :点击“查看数据包”,可对通过脱机载体移送到本目标平台分发数据包进行查询。并选择查询得到的分发数据包信息列表中的条目,打开右键菜单,点击“数据包导入”导入数据包中的数据到本目标平台数据库。

点击“查看导入数据”,可对已导入数据库的数据包内容进行查看。

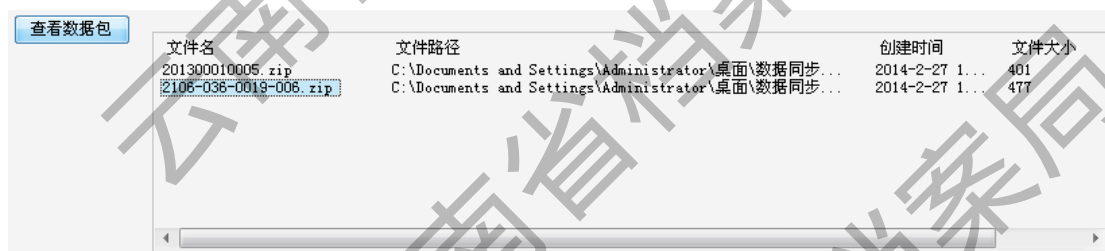


图 24 查询分发数据包信息列表

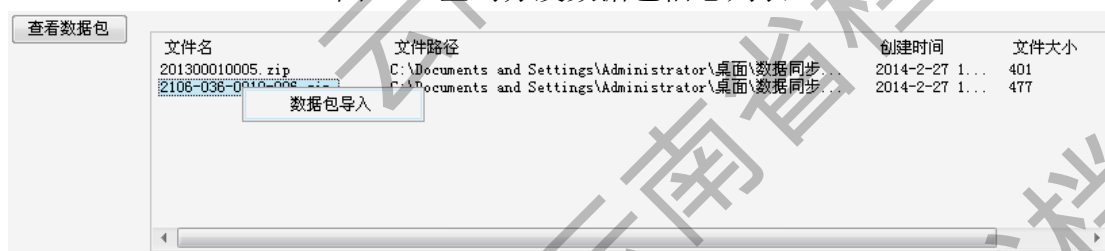


图 25 分发数据包导入



图 26 分发数据包导入成功后系统提示信息

查看导入数据	档号	题名	保存期限	年度	状态	是否有原文	创建时间	修改时间	完整性验证
	201300010002	数字档案馆建设指南	30年	2013	已生成确认...	有	2014-02-...	11	22
	201300010005	数字档案馆建设指南	30年	2013	已生成确认...	有	2014-02-...	11	22
	2106-036-...	云南省物资供应局革命委员会关于...	永久	1972	已生成确认...	有	2014-02-...	11	455927

图 27 查询导入数据

10.2.2.5 生成校验数据包

业务说明 :生成校验数据包功能将本次数据更新情况封装为确认数据包(包括档案唯一性标识、同步流向、同步完成标记、数据变化标记、档案数据完整性标记等等),并将此确认数据包移送到数据来源平台。

操作说明 :在导入数据查询结果列表中,选择数据条目,打开右键菜单,点击“生成确认数据包”进行数据包打包。

档号	题名	保存期限	年度	状态	是否有原文	创建时间	修改时间	完整性验证
201300010002	数字档案馆建设指南	30年	2013	已生成确认...	有	2014-02-...	11	22
201300010005	数字档案馆建设指南	30年	2013	已生成确认...	有	2014-02-...	11	22
2106-036-...	云南省物资供应局革命委员会关于...	永久	1972	已生成确认...	有	2014-02-...	11	455927

图 28 生成确认数据包

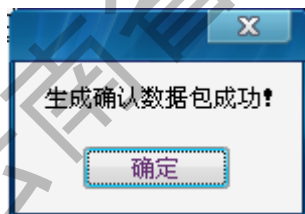


图 29 生成确认数据包成功后系统提示信息

10.2.2.6 确认数据包移动

业务说明 :使用人工方式通过脱机载体将生成在目标平台的确认数据包移动到来源平台。

操作说明 :拷贝目标平台“\bin\Release\send\”目录下的确认数据包“R 档号.zip”.移动到来源平台“\bin\Release\recv\”目录下。



图 30 移动确认数据包示意图

10.2.2.7 验证同步结果

业务说明 :读取目标平台通过脱机载体移送来的确认数据包 ,将确认数据包中的同步管理数据与来源平台的同步管理数据进行比对 ,如果比对结果一致 ,则修改数据库中的同步标识为已同步成功 ,如果比对结果不一致 ,则记录此数据 ,并重新准备同步流程。

操作说明 :在“来源平台”界面档案条目列表选择档案条目数据 ,打开右键菜单 ,点击“接收确认数据包” ,完成所选档案条目的目标平台确认数据包的读取、解析、对比、同步等 ,验证同步结果成功后 ,改变完成标识系统等信息 ,显示返回完整性校验等信息。

创建时间	修改时间	同步时间	同步流向	完成标识
2014-02-27 11:1...	2014-02-28 12:1...			
2014-02-28 09:2...		2014-2-28 9:27:17	A服务器	已打包
2014-02-28 11:4...				

- 数据打包
- 接收确认数据包
- 删除数据
- 选择模板

图 31 选择档案条目右键菜单



图 32 验证同步结果成功后系统提示信息

同步时间	同步流向	完成标识	数据变化标识	完整性校验	返回完整性校验
2014-2-28 9:27:17	A服务器	已确认同步	新增	857579	857579
2014-2-28 14:20:11	C服务器	已打包	新增	919506	

图 33 验证同步结果成功后列表显示信息

附录

云南省档案馆档案数据同步管理办法（试行）

一、 为规范云南省档案馆档案数据同步工作，确保档案数据的档案数据的真实、完整、可用和安全制定本办法。

二、 本办法所指的档案数据同步是指将完成整理、鉴定等工作的档案数据从局域网系统向屏蔽机房、政务网、互联网等系统传送保持系统间数据一致性的工作。

三、 云南省档案馆数据同步工作由备份管理处负责组织实施，处领导负责数据同步工作监督和检查，系统管理员负责数据同步的组织 and 确认，数据管理员负责数据同步的操作。

四、 数据同步采用单向方式，即从局域网系统向屏蔽机房、政务网、互联网的系统同步。

五、 局域网内数据发生新增、删改等变化后，应及时向屏蔽机房系统同步。

六、 向政务网、互联网同步数据应按数据管理规定，把经局馆领导同意后实施。

七、 数据同步前系统管理员应填写同步操作记录单，详细登记同步数据的内容、数量及同步目标系统，数据管理根据记录单进行同步操作并签字确认，同步操作完成后系统管理员应对同步结果进行签字确认。

八、 数据同步介质应按照脱机介质管理有关规定进行统一管理，登记编号并明确使用范围，原则上应专网专用。故障或者淘汰不用的介质应保密要求进行安全处理。

九、 档案数据同步可以使用光盘、硬盘作为介质，同步完成过后档案数据光盘应进行安全销毁，硬盘数据应进行安全删除。验证数据同步应使用一次性写入光盘作为介质，同步完成后验证数据光盘一般应保存三年以上。

十、 本办法自印发之日起实施，由云南省档案馆负责解释。

参考文献

图书：

- [1]李泽锋.基于 OAIS 电子文件管理系统体系研究[M].上海：上海世界图书出版公司，2010.
- [2]邵佩英.分布式数据库系统及其应用[M].北京：科学出版社，2005.
- [3] [美] Ramez Elmasri,Shamkant B. Navathe ，数据库系统基础（第6版）[M]. 李翔鹰[等]译. 北京：清华大学出版社，2011.
- [4] [美] Peter Rob,Carlos Coronel ，数据库系统设计、实现与管理（第6版）[M]. 张瑜，杨继萍[等]译. 北京：清华大学出版社，2005.
- [5][美]塔嫩鲍姆.分布式操作系统 (Distributed.Operating.Syst em). (美) [M]. 陆丽娜,伍卫国,刘隆国, [等]译. 北京：电子工业出版社，2008.
- [6][美]HectorGarcia-Molina,Jeffrey D.Ullman,Jennifer Widom. 数据库系统全书 [M]. 岳丽华，杨冬青，龚育昌，唐世渭，徐其钧[等]译. 北京：机械工业出版社，2003.
- [7] 朱小怡[等]. 数字档案馆建设理论与实践[M]. 张瑜，杨继萍[等]译. 上海：华东师范大学出版社，2007.

期刊：

- [1]钱毅. OAIS 对数字档案馆系统技术路线和管理策略的启

- 示[J]. 档案学研究,2009,04:46-49.
- [2]章燕华,刘霞. OAIS 参考模型:数字资源长期保存的概念框架[J]. 浙江档案,2007,03:38-42.
- [3]李明娟. OAIS 参考模型与数字信息长期保存[J]. 图书情报知识,2007,05:65-69.
- [4]吴霜,郑海涛. XML 在分布式数据同步中的应用研究[J]. 现代计算机(专业版),2004,08:31-34.
- [5]黄玉明,周建武,吴彬松. 安徽省电子文件中心设计与建设[J]. 中国档案,2009,08:11-13.
- [6]陶水龙. 大数据时代下数字档案馆面临的机遇与挑战[J]. 中国档案,2013,10:66-68.
- [7]蔡学美. 当前档案信息化建设的目标和主要任务[J]. 四川档案,2013,02:14-15.
- [8]蔡学美. 档案工作中使用“电子档案”概念的基础和作用[J]. 档案学研究,2011,01:4-6.
- [9]钱毅. 档案数据库的规范和质量控制[J]. 档案学通讯,2007,05:53-56.
- [10]钱毅. 档案数据库建设中存在的问题及解决思路[J]. 档案学通讯,2006,04:45-48.
- [11]黄玉明. 电子文件和相关数字资源封装案例比较[J]. 档案学研究,2012,05:55-58.
- [12]黄玉明. 电子文件(档案)管理系统建设的需求分析[J]. 中国档案,2011,02:34-36.

- [13] 黄玉明. 电子文件封装策略研究[J]. 中国档案,2010,04:65-67.
- [14] 黄玉明. 电子文件存档格式需求分析与战略研究[J]. 档案学通讯,2010,02:63-66.
- [15] 黄玉明. 电子文件格式困局破解之道[J]. 中国档案,2010,07:56-58.
- [16] 屠跃明,黄永文. 对 OAIS 参考模型的研究[J]. 档案学研究,2007,02:45-47.
- [17] 周峥嵘,王琤,何文春. 分布式气象元数据同步系统的探索研究[J]. 应用气象学报,2010,01:121-128.
- [18] 王力萌. 分布式社保计算机系统数据同步技术研究[J]. 光盘技术,2009,01:48-49.
- [19] 周伟,刘丰. 分布式数据库数据同步技术研究[J]. 计算机与信息技术,2010,10:13-16+20.
- [20] 周自斌,朱方洲. 分布式数据库的数据同步技术研究和应用[J]. 电脑学习,2004,02:22-23.
- [21] 丁鲲,严浩,刁兴春. 分布式数据库数据同步技术研究[J]. 海军工程大学学报,2004,05:100-104.
- [22] 崔伟,汪诗林. 分布式系统中数据同步机制的研究与实现[J]. 计算机工程与设计,2007,10:2259-2261.
- [23] 金松河,赵进超. 分布式异构数据库下的数据同步系统研究[J]. 云南民族大学学报(自然科学版),2009,02:162-164.
- [24] 张瑛,夏克俭,张法明,蒋晨晓. 分布式异构数据库数据同

步系统的研究与实现 [J]. 小型微型计算机系统,2007,10:1803-1806.

[25]胡凤华,袁继军. 高校数字档案馆信息资源整合交换的策略及应用[J]. 档案学研究,2011,01:43-46.

[26]李萍. 关于分布式异构数据库数据同步系统的研究[J]. 电大理工,2008,01:63-65.

[27]胡星火. 基于 OAIS 的数字信息长期保存研究[D].南京航空航天大学,2008.

[28]叶丽. 基于 OAIS 信息包的电子文件元数据封装策略研究[J]. 档案学研究,2007,04:54-57.

[29]张智雄,郭家义,吴振新,林颖. 基于 OAIS 的主要数字保存系统研究[J]. 现代图书情报技术,2005,11:1-9+13.

[30]卢宇,龚忠友,吴进营,苏伟达,朱丽,吴允平,蔡声镇. 基于 WEB 服务的分布式异构数据同步设计[J]. 微计算机应用,2011,12:47-53.

[31]郗海龙,杨延双. 基于 XML 的分布式异构数据库数据同步系统应用研究[J]. 沿海企业与科技,2009,02:38-40.

[32]沈敏,许华虎,季永华,万杰. 基于 XML 的分布式异构数据库数据同步系统的研究 [J]. 计算机工程与应用,2005,05:184-186.

[33]龙欣,张旭,席大春,周福平. 基于 XML 业务无关的分布式数据库数据同步策略 [J]. 华中科技大学学报(自然科学版),2003,05:37-40.

- [34]董立岩,毛锐,余宜诚,王利民,黄乐,殷涵. 基于分布式多服务系统的数据同步方法[J]. 吉林大学学报(理学版),2011,04:745-749.
- [35]黄晨晖,林泳琴. 基于分布式SQL和流复制的数据同步系统[J]. 微计算机信息,2010,36:266-267+124.
- [36]花胜强,王丽丽,朱志丹. 基于隔离和异构采集系统的多坝数据集成的研究[J]. 水电自动化与大坝监测,2013,01:54-57.
- [37]尹飞,熊政,祝永晋. 基于物理隔离技术的触发式数据库同步方案[J]. 江苏电机工程,2010,01:35-37.
- [38]曹静. 基于物理隔离的分布式数据库同步[J]. 计算机工程,2008,16:186-187+190.
- [39]蔡学美. 及时移交 规范接收——《电子档案移交与接收办法》发布[J]. 中国档案,2013,01:34-35.
- [40]胡红霞. 论数字档案馆的建设——数字档案馆档案信息采集的研究[J]. 中国教育研究论丛,2005,00:25-27.
- [41]潘晓秋. 美国数字资源长期保存主要成果及其启示[J]. 现代情报,2008,05:98-101.
- [42]李敏. 浅谈开放的档案信息系统(OAIS)[J]. 兰台世界,2007,16:20-21.
- [43]冯欣. 使用Java与XML在异构环境下建立分布式应用的实现[J]. 长春理工大学学报,2002,03:38-41.
- [44]张金英,李蜀瑜. 数字档案馆数据交换系统的设计与实现[J]. 计算机技术与发展,2012,11:100-104.

- [45]时照建. 数字档案馆低成本数据备份方法[J]. 兰台世界,2009,21:18.
- [46]徐震. 数字档案馆数据备份系统方案的规划[J]. 兰台世界,2006,18:21-22.
- [47]蔡学美. 数字时代档案原始性的特征和应对策略[J]. 中国档案,2012,07:68-70.
- [48]刘永亮,张卫红. 数字校园中的数据交换平台设计[J]. 中国教育网络,2011,Z1:91-93.
- [49]张晓林. 数字信息的长期保护问题[J]. 图书馆,2001,05:7-12.
- [50]程雪梅. 数字资源长期保存技术之探讨[J]. 图书馆理论与实践,2005,05:85-87.
- [51]商杏芬. 探讨内、外网物理隔离后数据交换及移动办公的解决方案[J]. 信息与电脑(理论版),2013,09:74-75.
- [52]蔡学美. 推进数字档案馆建设 提升各级档案馆信息化管理能力[J]. 中国档案,2011,03:56-57.
- [53]杨宏桥,吴飞,刘玉树,赵志云. 网络隔离与安全交换技术在 HIS 中的应用研究[J]. 医疗卫生装备,2008,02:45-47.
- [54]王艳丽. 我国档案界对 OAIS 参考模型的研究和应用[J]. 浙江档案,2009,06:26-28.
- [55]柯军,须文波. 物理隔离环境下数据安全转发与实现[J]. 微计算机信息,2006,15:64-66.
- [56]张震. 物理隔离技术分析及其数据安全转发模型[J]. 微

计算机应用,2004,01:32-36.

[57]张蒲生. 物理隔离环境下数据交换机制及安全性研究[J]. 微计算机应用,2002,04:242.

[58]刘亮超. 物理隔离——让数据河水不犯井水[J]. 互联网周刊,2001,13:107.

[59]南希 德罗梅迪(Nancy Deromedi) 本特利历史图书馆数字保管处负责人、助理研究馆员 周文泓 编译. 新一代档案馆的数字保管[N]. 中国档案报,2012-02-16003.

[60]张思达. 信息摆渡系统解决物理隔离下的信息交互问题[J]. 成都信息工程学院学报,2004,01:62-65.

[61]孙业毅,吴国凤. 一种基于物理隔离的数据库同步方案[J]. 计算机技术与发展,2007,11:175-178.

[62]唐利峰. 一种物理隔离数据交换系统的设计[J]. 信息技术,2009,10:30-32+35.

硕博士论文:

[1]周家生. 分布式网络数据同步技术研究与应用[D]. 哈尔滨工程大学,2011.

[2]谢金科. 银行监控系统的分布式数据同步研究及应用[D]. 西南交通大学,2012.

[3]姚超. 面向业务规则的分布式数据同步方法的研究[D]. 武汉理工大学,2013.

[4]王志强. 分布式应用中数据同步关键技术研究与应用[D].

石家庄铁道大学,2013.

[5]陈广明. 基于 SQLite 的分布式数据同步技术研究与应用[D].华南理工大学,2013.

[6]赵应钢. 异构分布式数据库数据同步系统设计与实现[D].华中科技大学,2007.

[7]余东. 基于 XML 和.NET 的分布式数据同步的研究与应用[D].合肥工业大学,2009.

[8]刘永毅. 异构分布式数据库远程数据同步的研究和设计[D].吉林大学,2010.

[9]戴婉荣. 基于分布式数据库的数据同步机制的研究与应用[D].武汉理工大学,2010.

[10]徐瑞雪. 基于 JAVA/XML 的分布式数据同步系统的设计和实现[D].大连海事大学,2011.

[11]赵亚陆. 异构分布式数据库中数据同步更新的研究与应用[D].大连理工大学,2000.

[12]杨源. 分布式数据库系统 D-SQL 中数据更新的设计与实现[D].电子科技大学,2001.

[13]温昱晖. 综合自动化系统中分布式数据库的应用[D].昆明理工大学,2002.

[14]蒲志林. 网络环境下分布式数据库同步机制的研究[D].大连海事大学,2003.

[15]张雄. 分布式数据库数据同步的研究与应用[D].华中科技大学,2006.

[16]陈锡彬. 物理隔离数据交换系统研究与实现[D].哈尔滨工程大学,2005.

[17]王琦. 电子办税服务系统中数据同步软件的设计与实现[D].南京理工大学,2011.

[18]周健. 数据同步技术在省级数据集中的应用[D].湖北大学,2012.

[19]罗冰. 实时在线双网隔离系统的设计与实现[D].南京理工大学,2010.

[20]毕建新. 数字档案馆存储研究[D].苏州大学,2004.

[21]解俞. 基于数字档案馆的电子政务数据中心模式研究[D].浙江大学,2007.

[22]孙琦. 数字档案馆的构建[D].吉林大学,2011.

其它:

[1]国际标准化组织 《Space data and information transfer systems — Open archival information system — Reference model》(ISO 14721:2003 (E)).

[2]国家档案局 《文书类电子文件元数据方案》(DA/T 46-2009) .

[3]国家档案局 《基于 XML 的电子文件封装规范》(DA/T 47-2009) .

[4]国家档案局 《版式电子文件长期保存格式需求》(DA/T 48-2010) .

[5] 国家质检总局 《电子文件归档与管理规范》
(GB/T18894-2002).

[6] 国家档案局 《数字档案馆建设指南》(2010).

[7] 国家档案局 《电子档案移交与接收办法》(2012).

云南省档案局
云南省档案局
云南省档案局
云南省档案局
云南省档案局