

《电子文件长期有效性保证的研究》 研究报告

安徽省档案局

2017年6月

国家档案局官网
WWW.SAAC.GOV.CN

目 录

摘 要:	3
1 绪论	4
1.1 课题的来源	4
1.2 课题研究背景	4
1.3 国内外研究状况	5
1.4 课题研究的对象、意义与价值	7
1.4.1 研究的对象	7
1.4.2 意义与价值	8
1.5 课题的研究方法	9
2 电子文件长期有效性的目标要求	10
2.1 有效性和有效期	10
2.2 目标要求	12
3 电子文件长期有效性保证基本策略	13
3.1 加工数据	14
3.2 定制环境	28
4 电子文件管理最佳实践——可信数字仓储	39
5 研究的主要成果、结论、创新之处和局限	42
5.1 主要成果	42
5.2 结论	43
5.3 课题的创新之处	44
5.4 课题的局限性	46
6 参考文献	47

国家档案局官网
WWW.SAAC.GOV.CN

4 电子文件管理最佳实践——可信数字仓储

实施数据加工和营造合适的存贮环境只是电子文件长期有效保存的具体策略，而这些策略的实施需要特定的组织形式，国际上，普遍认为：可信数字仓储（Trusted Digital Repositories, TDR）是实现电子文件长期有效性保障策略的最有效形式。有关 TDR 的进一步信息可以查看 RLG 和 OCLC 发表的《可信数字仓储：属性和职责（Trusted Digital Repositories: Attributes and Responsibilities）》报告，以及《可信数字仓储审核与认证：标准与审核表（Trustworthy Repositories Audit & Certification: Criteria and Checklist, TRAC）》。在我国现阶段电子文件长期有效性管理实践中，TDR 的基本形式是数字档案馆（如青岛市数字档案馆、本课题的第 2 个实践平台安徽省数字档案馆）和电子文件中心（如本课题第 1 个实践平台安徽省电子文件中心）。构建合格的 TDR 的基本要求可以归纳为以下几个方面：管理电子档案的主体——机构及人员，被管理的客体——电子档案信息资源，主体在管理客体过程中所开展的活动——业务，把主客体和业务活动联系在一起的环境——计算机系统，以及综合管理要素等。

（1）组织机构要求

应当具备长期保存并提供利用电子文件的合法职能，有

相应的编制、有能胜任工作的业务和技术人员、有比较稳定的经费来源、有相应的场所和开展业务所需的设施、设备，制订有相应的长期保存策略，电子文件管理流程清晰，具备归责机制等。

(2) 电子文件信息模型要求

包括本文提及的电子文件数据格式要求，如数据抽取与重构规范，格式分析、注册、转换机制，指定接收的电子文件格式类型、预定义的标准格式规范，元数据方案、著录要求，归档信息包的封装标准与 schema 定义，封装包校验规则、修订规则，递交包与分发包模型等。

(3) 信息系统功能模型要求

电子文件长期保存和利用系统建设的核心要求是必须遵从开放档案信息系统参考模型 OAIS。OAIS 定义的系统功能模型由 6 个功能实体构成，包括档案捕获 (Ingest)、档案存储 (Archival Storage)、数据管理 (Data Management)、档案利用 (Access)、系统管理 (Administration) 和，保存规划 (Preservation Planning)。有关 OAIS 的完整信息可以参见 ISO 14721^[17]。当然，TDR 所要求的除了系统功能之外，还有非功能要求，如系统的并行用户数量，并行事务处理能力，海量检索响应时间，电子文件仓储容量，数据库管理能力，可持续服务时间，宕机后系统恢复时间等，

[17] ISO 14721. Space data and information transfer systems — Open archival information system — Reference model[S]. 2003-03-01.

这些也是电子文件长期保存和利用系统建设中应当考虑的部分。

(4) 业务要求。业务要求是在 TDR 体系内，从业务流程、业务标准和业务参与者的角度来规范业务活动，确保电子文件的收、管、存、用等各项管理业务有条不紊，各岗位人员各司其职，TDR 正常运转。对于数字档案馆和电子文件中心来说，通常要结合实际，进行业务流程重构，尤其需要就影响电子文件长期有效性的核心节点，如电子文件归档、电子档案移交与接收环节建立有效机制。

(5) 管理要求。包括综合管理要素，如制度环境的准备，风险管理（如风险因素、特点的识别，风险管理规划、监控、评估，建立风险应对体系、完善应对措施等），变化管理（即具备在漫长的电子文件生命周期管理过程中根据需要作出改变的能力，如根据技术革新，调整技术路线，根据电子档案移交者的变化，调整移交策略等），以及日常管理和维护，审计、跟踪与问责等。

需要指出的是，电子文件长期有效性保障策略会随着科研和实践的发展而不断推陈出新，前述“加工数据”与“定制环境”的诸多策略仅仅是从理论上归纳的“超集”。电子文件长期保存机构在具体工作实践中不需要照单接收，逐个使用，要结合自身特点，在“超集”中选择若干策略，形成“套餐组合”，以此满足电子文件长期有效性保障的需求。

青山遮不住，毕竟东流去。电子文件海量生成与长期保存是历史的必然趋势。唯有与趋势为友，系统地、全面地分析与实施电子文件长期有效性保障策略，才能从容应对信息技术带给我们的严峻挑战。

5 研究的主要成果、结论、创新之处和局限

5.1 主要成果

1) 电子文件管理系列标准

《文书类电子文件元数据方案》(DA/T 46-2009)

《版式电子文件长期保存格式需求》(DA/T 47-2009)

《基于 XML 的电子文件封装规范》(DA/T 48-2009)

《电子文件元数据基本集（报批稿）》

2) 相关科研论文

安徽省电子文件中心建设的思路与做法

关于匹兹堡模式与哥伦比亚模式重新思考

安徽省电子文件中心设计与建设

电子文件封装策略研究

电子文件存档格式需求分析与战略研究

电子文件中心的定位与核心取向

电子文件格式困局破解之道

文件/档案及相关资源元数据再研究——国际源流与中国

体系构建

国家档案局官网
WWW.SAARC.GOV.CN

电子文件（档案）管理系统建设的需求分析
电子文件和相关数字资源封装案例比较
电子档案接收与典藏信息建模
档案信息化发展趋势探析
安徽探索有中国特色的电子文件管理之路
OAIS 与电子文件管理系统
安徽省电子文件中心存储系统建设方案
安徽省电子文件中心系统接收与长久保存电子文件的策略
基于 XML 的对象封装技术
3) 电子文件长期保存系统
安徽省电子文件中心
安徽省数字档案馆

5.2 结论

本课题采用两个“一揽子”解决方案，系统性地解决电子文件长期保存和利用（有效性）问题。

一是理论和实践“一揽子”。即采用：电子文件管理系列标准和系列论文+电子文件长期保存信息系统+电子文件中心（数字档案馆）三种理论和实践形式，系统性地提供电子文件长期有效性解决方案。课题重点研究和制订了《文书类电子文件元数据方案》、《版式电子文件长期保存格式需

求》、《基于 XML 的电子文件封装规范》等，对国内外相关理论和实践成果进行了比较全面的总结、借鉴，并创新应用，基于 OAIS 架构，研发了电子文件长期保存系统，在省级层面建设了满足电子文件长期保存和利用需求的电子文件中心（数字档案馆）。

二是技术策略和方法的“一揽子”。即：采用对电子文件进行处理（通过数据抽取与重构、数据风干、格式分析与转换、格式注册与开放、数据规范化、版式化、元数据定义、封装、数字技术校验、“四性”检测等方法进行加工数据）和营造合适的保存环境（通过技术典藏、仿真、介质更新、迁移、定义数据物理结构、资源标识、异构数据交互、载体选择、信息安全控制、备份、容灾、数据冗余、云存储、互操作）的策略、方法，系统性地提供电子文件长期有效性解决方案。课题结合我国“档情”，在上述策略和方法中选择数据规范化、元数据定义、封装、数字技术校验、“四性”检测、资源标识、载体选择、信息安全控制、备份、容灾、数据冗余等部分策略和方法，同时结合电子文件保存的物理的构建，来保证电子文件存储载体可读、计算机可正确解析、人可理解、可原貌呈现，做到电子文件的长期有效。

5.3 课题的创新之处

本项目成果为标准、论文、系统和最佳实践，在国际和

国内同类项目中处于先进水平。《文书类电子文件元数据方案》是 ISO23081 发布后，继澳大利亚、加拿大、欧盟之后发布的第 4 个国际同类标准。《版式电子文件长期保存格式需求》是国际上第一个版式电子文件存档格式需求标准。是“核高基”项目《版式文档标准制定与软件产品研发及产业化》(2010ZX01044-002) 的基点，是我国自主知识产权的版式文档格式 OFD 的基础。《基于 XML 的电子文件封装规范》是继美国元数据编码和传输标准 (METS)、澳大利亚维多利亚电子文件策略 (VERS) 后发布的第 3 个国际数字资源封装标准，是 ISO23081 发布后，第 1 个电子文件封装标准，同时也是符合 ISO3081 元数据框架和类型的第 1 个标准。安徽省电子文件中心是我国首个省级电子文件中心，是首个按照 OAIS 功能架构和信息模型框架结构，实现对电子文件及其元数据进行全过程、全方位管理的系统，是首个可以进行 EEP 封装和对 EEP 文件进行管理的电子文件管理系统。

本项目为应对信息时代电子文件管理给档案工作带来的挑战，探索解决当前电子文件管理存在的诸如：信息割据、管理失范、流向不明、缺乏保障、没有接轨等问题，从电子文件的信息模型、电子文件管理系统的功能模型和业务模型等方面开展研究。采用了国际最先进的电子文件管理理论和方法，通过制订标准，建立信息模型，对元数据和内容文档及签名块进行封装、规范电子文件数据长期保存格式等技术

手段，从根本上保证了电子文件的长期有效性，使电子文件彻底摆脱了对特定软硬件的依赖。安徽省电子文件中心建设，一是创建了电子文件中心应用系统功能模型，二是根据业务需求和管理实际，创建了业务模型。作为项目成果的3个标准三位一体，与信息管理系统建设和电子文件中心建设紧密结合，理论联系实际，从标准到系统、从理论到实践，在国内率先对电子文件长期有效性保证进行系统性研究，同时也开创了我国电子文件规范化管理的先河，三个标准跟踪国际前沿，等高对接，其中《版式电子文件长期保存格式需求》为国际首创，安徽省电子文件中心为国内首家省级电子文件中心，被《中国档案报》誉为“破冰之举”，人民日报以《守护国家的数字记忆》予以报道。安徽省经济和信息化委员将项目以“省档案局探索出保证数字资源真实、完整、有效的方法和途径”为题，评选为2012年度安徽省信息化十件大事。经国家档案局建议，安徽省委省政府给予专项表彰。国家档案局指出“三个标准借鉴国际先进做法，并充分考虑我国电子文件管理的实际制定完成的，其内容达到国内领先水平”，“为全国档案事业的科学全面发展提供了先进的经验和做法，发挥了引领和示范作用。”

5.4 课题的局限性

课题主要研究方向为电子文件管理的业务和技术，较少

涉及电子文件管理的体制和机制，同时在电子文件种类方面较多注重文书类电子文件，可以为其他类电子文件的管理提供参考。

6 参考文献

ISO 15489 信息与文献 文件管理

ISO/PDTR 23081 《信息与文献-文件管理流程—文件元数据

ISO-14721 《开放档案信息系统参考模型 (OAIS)》

ISO 11179 信息技术 元数据注册系统 (MDR)

ISAD(G): General International Standard Archival Description(Second Edition)

ISDF: International Standard for Describing Functions (First Edition)

ISDIAH: International Standard for Describing Institutions with Archival Holdings (First edition)

PREMIS Data Dictionary for Preservation Metadata version 2.0

VERS Metadata Scheme PROS 99/007 (Version 2)
Australian Government Implementation Manual: AGLS Metadata version 2.0

Government of Canada Records Management Metadata

Standard

APPENDIX 9 TO THE MoReq2 SPECIFICATION: METADATA
MODEL

ISO/PDTR 26102. 3 Records Management—Requirements
for long term preservation of electronic records

ISO 19005-1:2005 Document management -
Electronic document file format for long-term
preservation - Part 1: Use of PDF 1.4 (PDF/A-1)

ISO/TR 18492:2005(E) Long-term preservation of
electronic document-based information

Evaluating File Formats for Long-term Preservation

GB / T 9705 文书档案案卷格式

GB/T 15418 档案分类标引规则

GB/T 16964. 1-1997 信息技术字型信息交换第 1 部分：
体系结构

GB/T 18793-2002 信息技术可扩展置标语言（XML）
1.0（W3C RFC-xml:1998，NEQ）

GB/T 18894 电子文件归档与管理规范

GB/T 19488. 2-2008 电子政务数据元 第 2 部分
数据目录

GB/T 19667. 6-2006 基于 XML 的电子公文格式规范 第 6
部分 归档（草案）

GB/T 20163-2006 中国档案机读目录格式

GB/T 21063.3-2007 政务信息资源目录体系 第3部分
核心元数据

GB/T 29194-2012 电子文件管理系统通用功能需求

GB / T 31913-2015 文书类电子文件形成办理系统通用
功能要求

GB / T 31914-2015 电子文件管理系统建设指南

GB / Z 23283-2009 基于文件的电子信息的长期保存

DA/T 18-1999 档案著录规则

DA/T 22-1999 归档文件整理规则

EJ / T 1224-2008 核电电子文件元数据

数字档案馆建设指南（档办〔2010〕116号）

数字档案室建设指南（档发〔2014〕4号）

国家档案局官网
www.saac.gov.cn