SHANGHAI UNIVERSITY

毕业设计（论文）

**UNDERGRADUATE PROJECT (THESIS)**

**题 目： 论电子文件前端控制技术与方法**

**学 院 图书情报档案系**

**专 业 档案学**

**学 号 14170175**

**学生姓名 钱磊**

**指导教师 金波**

**起讫日期 2016.12.27-2017.4.26**

论电子文件前端控制技术与方法

目录

[摘要 2](#_Toc480958217)

[ABSTRACT 2](#_Toc480958218)

[一、绪论 3](#_Toc480958219)

[1．1选题意义 3](#_Toc480958220)

[1．2研究综述 3](#_Toc480958221)

[1．3研究思路和方法 4](#_Toc480958222)

[1．4主要创新和贡献 4](#_Toc480958223)

[二、电子文件的形成和运行 5](#_Toc480958224)

[2．1电子文件的特点 5](#_Toc480958225)

[2．2电子文件的作用 6](#_Toc480958226)

[2．3电子文件的形成规律及生命周期 6](#_Toc480958227)

[三、前端控制理论 7](#_Toc480958228)

[3．1前端控制理论的提出 7](#_Toc480958229)

[3．2前端控制理论的产生背景 7](#_Toc480958230)

[3．3前端控制理论的内容和意义 8](#_Toc480958231)

[四、电子文件的前端控制 9](#_Toc480958232)

[4．1电子文件的双套制问题及对策 9](#_Toc480958233)

[4．2电子文件运行过程的一致性问题及对策 10](#_Toc480958234)

[4．3电子文件信息的连续性问题及对策 11](#_Toc480958235)

[4．4电子文件的知识化利用问题及对策 11](#_Toc480958236)

[4．5 电子文件前端控制的核心 12](#_Toc480958237)

[五、前端控制技术和方法 13](#_Toc480958238)

[5．1基于云计算实现批量文件前端控制 13](#_Toc480958239)

[5．2利用MD5信息摘要算法确保文件完整性 15](#_Toc480958240)

[5．3采用新型数字签名算法确保文件真实性 16](#_Toc480958241)

[5．4为电子文件添加语义性内容增强可读性 18](#_Toc480958242)

[六、有关前端控制系统的建议 19](#_Toc480958243)

[6．1电子文件前端控制系统技术上的建议 19](#_Toc480958244)

[6．2电子文件前端控制系统管理上的建议 20](#_Toc480958245)

[七、前端控制应用案例 20](#_Toc480958246)

[7．1 HTML5 Canvas API图片处理与前端控制 21](#_Toc480958247)

[7．2 Web Audio API音乐可视化与前端控制 21](#_Toc480958248)

[7．3 通过火狐浏览器查看文件语义化带来的可读性 22](#_Toc480958249)

[致谢 22](#_Toc480958250)

[参考文献 23](#_Toc480958251)

[附录 23](#_Toc480958252)

# 摘要

在电子文件数量呈指数型增长的情况下，面向电子文件的前端控制思想为电子文件档案工作的管理方式和管理时机提出了新的要求，同时落后的管理水平和技术的快速发展让我们不得不重新思考电子文件的管理模式，才能在海量的电子文件中，对其进行精选甄别，做到取其精华，去其糟粕，真正做到重要电子文件的有效管理和确保其档案价值的实现。

基于电子文件特性和前端控制理论，本文立足当下前端控制痛点问题，提出解决对策并在此之上充分思考和利用计算机的相关原理和技术，分析各自对前端控制的利用可行性，并通过开发实例从视觉上展现前端控制的具体细节。最终对集成可行原理和技术的电子文件前端控制系统提出了相关建议。

**关键词：**前端控制 技术方法 前端系统

# ABSTRACT

In the case of an exponential growth in the number of electronic documents, the front-end control of electronic documents for the electronic file archives management and management of the timing of the new requirements, while the backward management level and the rapid development of technology so that we have Rethinking the management mode of electronic documents, in order to select the quality of electronic documents, to take its essence, to its dregs, truly important management of electronic documents and to ensure the realization of its archival value.

Based on the characteristics of electronic file and front-end control theory, this paper based on the current front-end control pain point problem, put forward solutions and on the full consideration and use of computer-related principles and techniques to analyze their use of front-end control feasibility. From the visual display of the specific details of front-end control. And finally put forward the relevant suggestions on the electronic file front-end control system which integrates the feasible principle and technology.

**Keywords:** Front end control Technical method Front end system

# 一、绪论

# 1．1选题意义

21世纪以来，信息技术成为高科技的强大引擎已成为不可否认的事实，以信息技术为核心的跨学科发展轨迹也是纵横交错，同时置身于档案学科，有别于传统档案的数字档案资源日益丰富，改变了以往档案信息资源的存在形式和形态，电子文件的数量持续激增，同时档案数字化步伐的加速以及深入社会各方面的触角不断延伸，让传统档案数字化后的文件以及产生于电子环境的归档电子文件的管理显得更为重要和突出，而前端控制思想根据文件生命周期理论和全程控制原则，对电子文件从形成、产生到归档的整个过程进行统一规划和要求，实现对电子文件的全程管理和超前控制。

对于电子环境下的前端控制研究不仅是档案学科自身发展所面临的问题，也是档案学科自身对于当下外部环境所展现出的应激性措施以及完善学科内涵的需要。电子文件的前端控制对于确保电子文件的真实性、可靠性、可读性将起到重要作用。而对于电子文件的前端控制研究将有利于实现档案工作前置，有利于实现归档前后档案工作重心的平衡，以及为电子文件的前端控制提供科学的实践经验和指明一条清晰的道路。

## 1．2研究综述

本文除了绪论篇以外分为六个部分，分别从电子文件的形成和运行介绍了电子文件的特点、作用、形成规律及生命周期从而深入探讨电子文件有别于其它文件的核心内涵，其中不乏较为理论性的知识，但是为了确保研究课题的内容的完整性、可读性、可理解性，本文在探讨电子文件的相关内容的同时增添了一些丰富的话题，旨在研究电子文件基础理论的同时确保相关内容与时俱进、具有可研究性。

第三章，在紧扣第一章有关电子文件的相关背景知识、核心内涵等基础上引出了前端控制理论，尽管前端控制思想在传统纸质文件的管理实践活动中就已经早有显现，但确是在现代电子文件的管理中彰显其思想的重要性并在现代文件管理活动的实践当中起到中流砥柱的指导作用。该章从前端控制理论的来源解析了前端控制的核心思想，让我们深刻感受到前端控制思想在档案管理活动中的先进性、前瞻性、务实性。最后也从思想的价值和作用深入探讨电子环境下前端控制的重要意义以及对档案工作质量的影响。

第四章，即进入了本文的重要章节，众所周知，前端控制在电子环境下的实施，一直表现着管理与技术的系统复杂性，具体表现的问题有，电子文件运动特性和静态属性冲突，局部属性和全局属性的冲突等等。简单而言，即双套制带来的问题。本质是技术发展与管理理论研究的失衡。该章以系统的视角分析了电子环境下前端控制面临的问题以及产生的原因，考虑到技术体系的结构特征，如一致性、层次性等，基于元数据捕获、管理提出了一些面向公共元数据的结构化、层次化对策以及对所提问题的综合解决对策。

第五章，正如我们所知道的，电子文件学科是一门应用性很强的学科，对技术展现出相当大的依赖性，而当下以计算机技术为核心的信息技术革命的浪潮正愈演愈烈，跨学科发展也让学科联系更为密切、相互知识迁移更为频繁，所以该章基于当下的重要技术、算法和原理，为电子环境的前端控制提供了具体的技术解决方案，不仅阐述了相关技术背景，更让实践思想深入研究成果。

第六章，分别以技术和管理视角对电子文件前端控制系统提出了一些实践和规范上的建议和看法。有利于在业界保持电子文件前端控制系统按照统一规范运行运作以及维护以达到技术上的互操作性和管理上的协调一致，有利于深入贯彻前端控制思想并让电子文件前端控制系统发挥重要作用。

第七章，作为结尾部分，笔者开发了几个简单实例并对相关内容作简要说明，在基于Web新型技术的基础上，利用上述研究成果，从视觉和分析两个方面依次展现了MD5算法确保电子文件的完整性、真实性的利用研究成果以及语义化在增强电子文件可读性上的利用研究成果，再次升华前端控制思想对于电子文件的完整性、真实性、可读性所起到的重要作用和价值，也是理论研究和计算机实践相结合的一次整体呈现。

## 1．3研究思路和方法

本文主要研究面向电子文件的前端控制，基于主次分析，确定电子文件基础特性、前端控制理论为辅助研究，电子文件的前端控制技术和方法为主要研究对象。明确了主次关系将有利于文章核心的把握和笔墨自愿的配置。

在研究电子文件前端控制的具体细节之前，必须抓住电子文件的核心特性，以及前端控制思想的主要内涵，才能辨别二者对外接口，掌握电子文件核心特性对前端控制提出的具体要求，同时确定前端控制具体实施中所需要考虑的电子文件特性。并基于当下电子文件前端控制中出现的棘手问题，提出纲领性解决对策，明晰前端控制核心。从前端控制核心着手，利用先用原理和技术，基于类比的研究方法探讨前端控制细节，从而让研究内容充实而并不显得冗余。

最后基于理论和实践相结合的研究方法，通过开发实例应用理论研究成果，再次升华电子文件前端控制主题。

## 1．4主要创新和贡献

本文的主要创新和贡献之处在于把握电子文件基础特性和前端控制理论核心之后，深入探讨技术细节，在理论指导下，探讨计算机相关算法和技术原理在电子文件前端控制中的实现方法。并通过类比模拟电子文件运行过程中的相关对象和行为，验证技术可行性。真正为前端控制如何在实施中确保电子文件真实性、完整性、可读性提供了理论研究成果和技术实现，同时为电子文件前端控制系统的运作提出管理和技术上的建议。

# 二、电子文件的形成和运行

既然面向电子文件实现前端控制，对于电子文件的形成和运行做到充分了解和把握，将有利于前端控制实施中有的放矢，并在管理上控制成本效益关系。同时，当前端控制实施中发生问题的时候，也可以利用反向思维根据电子文件特性找出问题根源。

## 2．1电子文件的特点

信息的非人工识读性

存在电脑的各种电子文件信息本质上是一系列二进制位串，比如说字符文件，通过字符编码算法，能将字符文件底层的二进制位同各种字符映射起来，这样一来电子文件就具有信息的非人工识读性，就拿这个例子来说，它需要我们拥有一套字符编码才能让我们将二进制位翻译成人类可读的或者具有价值的信息，现代信息技术发展在给人类带来便利的同时也让信息和人类的关系发生了某种疏远。对于信息的非人工识读性，可以从以下几个方面理解。

一是信息物理实体呈现的自然规律信息是有别于经特定翻译形成的信息。有些人可能会说即使是一系列二进制串我也能读出他们的规律，也能识读出特定的信息。这是因为信息具有多种表达形式，而某种特定的信息表达方式从本身上就能表达出一些规律或特征，但这和信息物理实体经过某种映射或翻译表达出的信息是截然不同的，一定不能混淆。就拿二进制位来说，有时候它可能就代表一个数值大小，有时候经过特定翻译它也能代表图片的一个像素信息，但是从物理实体上它就是多个0和1的组合。这种组合固然能呈现出规律信息，但这种信息和我们要存储的，要识读出的信息的价值是有所区别的。

二是同样的信息位在不同的上下文环境中是具备不同信息含义的。而信息物理实体本身的规律信息未经过特定翻译是单一的。电子文件的信息非人工识读性是某个特定时间空间上的特性，当二进制位信息存储在载体上的时候，未通过媒介，比如说显示屏，音响等的呈现，是我们看不见的电路信息的形态，它是不能够被识读的，在未被读取到内存进行处理的时候，它也是不能够被识读的。当我们需要识读出它所要表达出的信息的时候，就需要将它调取到内存中进行特定于上下文环境的翻译，比如说显卡处理后形成图像图形信息，经过音频处理器处理后，形成音频信息，比如美妙的音乐。

三是信息的非人工识读性不仅是静态的特点也是动态的特点。信息不仅仅是存在于载体上，也就是二进制位时所表示的不可识读性，当经过特定翻译时，因为翻译算法或上下文环境必要因子的缺失也会导致二进制位的翻译失败，从而展现出信息的非人工识读性，比如说乱码，就是因为错误的编码算法导致比特和符号映射出错。这就是它不可识读性所呈现的动态特点。

系统的依赖性

如上述第三点，信息非人工识读性所呈现出的动态特性，让信息翻译过程及其依赖于上下文环境，而这个上下文环境就是包含多种事物集合的系统，如特定的处理器、特定的版本、特定的编码算法。一个某种上下文环境的信息要是在另一个上下文环境中使用的时候，必须考虑到上下文环境的兼容性，这也构成了系统之间的兼容性，当生成文件的系统以及操作系统以及硬件的更新换代与文件的生成系统不兼容时，我们需要对原有的生成系统进行记忆或者新的系统能够兼容原有系统或者有能力恢复到原有生产环境。以确保电子文件信息能够独立于系统环境被利用，即使这种独立性是人为提供的，确保原有系统生成的电子文件的可管理性、可读性。

信息与特定记录载体之间的可分离性

存储媒介的多样性和特定存储媒介记录信息的差异性导致了信息可以存储于多种媒介，也可以以多种形式被存储。大多数时候，信息不需要对特定载体从一而终，不离不弃。从纵向上，它可以存储在地址集合中的任意一块子集合，前提是具备权限。或根据需要随时扩展、缩小或改变其存储空间。横向上，它可以因为资源请求、信息转移等存储在多种设备、多个客户端上。如服务器集群、互联网资源请求、缓存等。信息与载体的可分离性使得电子文件不再具备物理意义上的实体状态，成为人们所指的形象上的虚拟文件、非实体文件。

除上述特点以外，从中也很容易知道信息还具有可变性、信息存储高密度性、信息的可操作性等等。

## 2．2电子文件的作用

电子文件的很多特点让电子文件有着不同于传统文件的诸多特征，其作用大大的改变了人们传统的生活方式、工作方式以及管理方式。比如之前由交大创业团队设计的一款流程自动化系统，就是实现文件货物化、以管道的方式流转于各个节点，然后各个节点负责具体的文件操作，从而使得文件传输和利用相互分离，相互解耦。大大促进了办公效率和改变了办公方式。各个节点均处于一个信息系统框架下，从而让现实中的各个档案管理节点、文件中心、机关档案室有效被虚拟化，并存在于同一个系统的控制中，有利于文档一体化管理工作的实现。从互联网角度来看，信息共享也会促成电子文件资源的共享。和以往实体共享方式比较，共享更加便利和频繁。电子文件可以集成多媒体信息，从而让归档电子文件形成的档案的记录内容更加丰富，表现力更为十足。基于电子文件的子系统的不断发展和集成，更加有利于国家综合信息系统的建设和发展。

## 2．3电子文件的形成规律及生命周期

对电子文件形成规律的把握，有利于理清文件的过程特点，抓住文件过程间的联系，从而能将前端控制粒度细化到与文件形成过程相对应的程度，实现更精细化控制。对电子文件生命周期的探讨，让我们能够从全局视角掌握电子文件的整个过程，实现前端控制内容的全局策划和安排，做到饱和控制和防止物力人力冗余浪费。

电子文件的形成规律是其从制作到生成的整个过程中所展现出的内在规律，但这种规律有别于传统文件的形成规律，因为电子文件生成过程中不仅具备自己的特点，其生成过程还受到相关业务人员、相关活动、文件系统以及电子文件管理系统的影响，因而其规律是整个活动的所体现的规律，认识到这点将防止在制定前端控制政策时，忽略相关控制对象的控制工作。

总的来说，电子文件的形成具有几个特点：电子文件的形成是与业务活动同步进行的，这是一般规律。电子文件的形成依赖于特定的系统。电子文件形成过程中信息和存储可以分离。电子文件的数据类型与结构呈现多样化。电子文件形成具有时间和空间上的双重动态性。对于这些基本特点的把握，能为前端控制的精细化提供具体内容和方向，也有利于对前端控制产生的问题提供解决依据。

文件生命周期理论的基本内容就是，文件从生成到销毁或者永久保存是一个完整的运动过程，这一个过程依据文件价值形态变化可以划分若干个阶段；文件在每一级段因为其特定的价值形态而与服务对象、保存场所和管理形式之间存在一定对应关系。

# 三、前端控制理论

对电子文件施加前端控制的理论指导或思想指导就是前端控制理论和思想，充分把握前端控制理论思想的核心内涵，将有助于总结面向电子文件的具体控制细节，让前端控制理论控制主要的、忽略次要的。从而有利于控制成本效益关系。

## 3．1前端控制理论的提出

所谓的前端控制，即根据文件的生命周期理论以及全程控制原则，对电子文件的整个生命过程进行统一的规划和要求，即生成到归档的过程加以控制，比如将预先设计的监督功能、档案管理功能以及其它符合设计要求的各个子功能嵌入到文件管理系统，如上文所说的管道化方式工作的、基于节点构建的流程自动化管理系统，从而实现档案工作业务流程的重建和与信息系统各个节点的映射，最终实现电子文件的超前管理和全局控制。

传统档案管理方法面对电子文件所展现的不自信表现的突出，也就是有些地方不能一味的利用传统管理方法，法国诺加雷认为档案工作应该重新审视档案管理环节，重新考虑文件生命周期对于文件的干预时机。传统管理中，文件只有在归档后，档案部门才对其进行实质性的控制，归档之前的工作由文书部门负责。而电子文件较传统文件表现出的特性，使得这种管理就成为问题。

在深入研究国际档案界管理理论之后，我国提炼出了前端控制理论，最早体现于上个世纪80年代的文书和档案工作，具体表现为术语的产生和工作的实践，前端控制理论的探索和发展，给文书工作和档案工作带来了新的机遇和挑战。

## 3．2前端控制理论的产生背景

电子文件前端控制工作主要是研究特定的文件类型，推动文件格式标准化，综合考虑各种文件类型，提高文件生存能力，互操作能力，以及促进利用、共享并同时确保文件的可读性、真实性、完整性，在这个基础之上，尽可能实现低成本原则，如降低系统兼容性处理成本，减少传输比特流大小，应用先进压缩算法。虽然单个文件不能体现这些前端控制细节的价值，但是基于文件的惊人数量，一定会放大前端控制的价值。文件的增长趋势呈指数型。而如此大量的文件，如不加以控制和甄别，即使技术允许，现行的一些档案标准也难以接纳不加控制的大批量文件。比如管理成本。对于这些质量参差不齐的文件，为了确保凭证性，也要施加控制。

电子文件的动态性、流动性让电子文件的安全问题比传统的文件显得更加突出。本质是因为文件的信息和载体可分离性，信息内容易改性。这种安全问题一旦发生，具有不可控性、扩散性，让安全问题的影响更为严重。从信息内容而言，文件信息容易丢失，容易成为死文件，从而造成具有重要价值的记忆的流失。

当下文件管理系统的现状也着实让人担忧。双套制管理制度的存在，文件分散和存在于各种载体以及备份的管理方式。很容易看出人们对于文件以及文件管理系统表示的不信任。因此要加强电子文件的管理和缓解人们对电子文件的担心，就必须加强电子文件的前端控制。

## 3．3前端控制理论的内容和意义

前端控制理论以档案管理的视角来看，说到底就是将档案管理的介入时机切换到或者提前到档案的形成文件的控制上。在电子环境上，具体体现就是对于电子管理系统的设计，无论是表单字段值得设计还是系统架构的设计，甚至采用的技术标准都要通通将有利于档案控制的原则、思想元素给融入到设计、架构、技术中。让档案思想相伴相随，而不是等到要归档的时候才加入档案管理要素。良好的设计、架构优良的、技术成熟稳定的电子文件管理系统将为理想的前端控制铺下良好的基础。

在优秀电子管理系统的基础上，前端控制主要体现在动态控制上，对于电子文件的生成、处理、传递、保管、利用等各个环节确保电子文件的可靠、真实、可读、可利用。具体内容包括：

生成方面，保证文件本身的内容、结构、背景或元数据等构成文件的必要要素的完整不可缺失。内容即是数据的承载容器，这是确保真实性、完整性的主要区域。但不是说其它方面就是不重要的，这是相对而言。结构提现出文件的数据集合方式，良好的文件结构得益于良好的生成系统，也便于文件的序列化和反序列化，从而实现文件的传输和利用以及持久化。

处理方面，加强文件处理事务的原子性。原子性，就是事务的一系列操作保持统一和兼顾，一旦其中存在某个操作失败，该事务就回滚到和未发生该事务一样，对于操作的文件而言，其状态在此情形下不会发生改变，防止文件的数据污染或文件的构成要素的组合方式发生改变从而形成脏数据、脏文件。

传递方面，前端控制内容主要要求有，在数据通路中，文件的要素形式组合可以被调整以至于加速传输或者减少带款，但是在传递预处理中，如内容编码、分块传输编码、压缩算法一定是可以逆向恢复的，甚至对于一些重要文件，损失几个字节的信息都是严重的。

保管方面，前端控制要求为了确保保管后再利用的可能性，对于保管的编码方式要前后统一，比如说你采用ASCII编码保存，那么你的文件中一定不能有中文字符，这时候如果需要保存可以采用UTF8编码方式保存，防止后续利用出现乱码问题。

利用方面，对于文件权限一定得施加控制，可读还是可写、可写是基于拷贝还是原文件。防止对于文件的真实性、完整性、可读性因为利用而造成破坏。

对于电子文件的前端控制，基于计算机科学软件工程，国际档案理事会提出的三阶段理论可用于知道电子文件管理系统有关生成、处理、传递、保管和利用文件技术细节的指导。

前端控制理论的价值在于能够确保电子文件的真实可靠性、完整性、安全性、可读性，为档案管理从摇篮中准备。也是技术开发中良好的电子文件管理系统的设计指导原则，标准和规范。即使对于非归档文件，也能促成电子文件的科学、合理、有效的管理，降低管理成本，实现管理效益等等。

# 四、电子文件的前端控制

电子文件的快速增长，使得前端控制成为必然。对于电子文件的处理，一般采用集中批量的方式进行处理，反之，逐件处理会造成大量的物力人力消耗，难以平衡成本和效益。这样一来，集中批量处理较逐件处理而言，本来就是在一定程度上牺牲了逐件处理具有的低出错性，对于文件质量就有更高的要求，更要求其出淤泥而不染，文件本身的稳定性、完整性、真实性、可读性等将是前端控制的焦点。立足于当下，对于电子文件的前端控制，有以下几点问题：

## 4．1电子文件的双套制问题及对策

所谓双套制，就是采取两套机制实现文件的持久化和利用，以档案视角就是一方面将电子文件以传统档案载体的形式进行转化和归档处理，另一方面数字化处理后所形成的电子文件作为传统档案的副本。

起初，这种方式对于文件数量不是很多，档案业务不是很繁杂的情况下，还是表现出符合期望的效果和良好的成本效益，但是随着文件数量的激增，档案各类业务的增多，双套制渐渐地表现出弊端，具体体现在，档案库房容量和归档文件数量之间的失衡，传统档案数字化成本的难以控制，物力和人力的大量消耗使得双套制显得心有余而力不足，传统的管理视角，以新的管理方式去填补双套制管理方式的不足甚至在一定程度、一定范围内取代双套制，当然后者是基于新管理方式的绝对优越性。

但是我们不能否认的是在管理水平和技术、生产力失衡的情况下，双套制在作为过渡性措施向单套制过渡的过程中，发挥了重要作用，而且随着管理水平和技术的不断发展，双套制一定给我们管理的改进提供入口方向和大量的历史借鉴以及经验的汲取。比如说对于双套制来说，并不是所有的电子文件都能提供等同的传统载体档案作为副本，如特定格式的电子文件，如三维地图、工程图等。反之，也不是所有的传统载体档案都能够数字化成电子文件的形式，如已破损的档案、已玷污的档案内容、即使能够数字化成电子文件副本，也可能会因为元数据的完整性、真实性、可靠性、可读性的缺失或难以把握造成电子文件副本即使存在但也难以传播和利用，难以实现档案管理的归宿。

前端控制为平衡管理和技术的发展失衡，结合相对滞后的管理水平和发展迅猛的技术提供了一种可行的解决方案。与传统各个管理环节对应的技术解决方案必须相互协调，实现有效接口，才能做到协同，促进双套制带来的问题的有效解决。

## 4．2电子文件运行过程的一致性问题及对策

电子文件生命周期内各个过程都有特定的工作内容和工作目标，在实施前端控制之前，其每个阶段所呈现的特点和档案属性的差异还是比较明显的，作用也各不相同。在没有实施前端控制的时候，这种差异性具体体现在两个方面，一是客体差异性是指文件各个阶段缺乏或者不完全具备的基本档案属性，如档案的信息属性，档案信息可以理解成消息、情报、知识、数据、资料的总称。在文件的生成阶段，还未经过分析是不具备情报价值的。另一个就是主体差异性，是相对于不同管理经验以及不同知识结构的人而言，文件各个阶段的属性与档案属性之间除基本属性以外的区别。

前端控制在电子文件的形成和运行以及利用等各个阶段中的工作就是要注入各种必要的档案属性，以此才能彰显前端控制思想贯彻于电子文件的整个生命周期，做到超前控制。如前文中所说的差异性，总体上来看，差异性越大，前端控制工作内容就越多，具体体现在客体差异性越多，电子文件要捕获的元数据要求就需要越完整，在电子文件管理系统中体现的就是文件的属性字段越多。主体差异性越大，对于前端控制方案的选定，可靠性评估等一系列行政过程就会延长前端控制时机，对于物力和人力资源的调动耗时都会影响到电子文件的前端控制。相对而言，前端控制面向客体差异性所要做的工作还是较为简单的，关键在于前端控制面向主体差异性所要做的工作。

无论是生命周期理论和连续体理论，电子文件从形成到归档或者销毁的连续过程中，都有特定内容和目的，但此时施加前端控制，不停注入档案属性的工作显然会与文件原有的过程产生交叉或者并行操作，这个时候极易引起各个方面的一致性问题。如元数据捕获工作与文件的形成、流通、利用之间的冲突。有些档案属性是基于多个归档文件总和所表现的属性，如档号的分配。那么对于构成的每一份文件都需要搜集其元数据内容，并确保较自身而言元数据内容的可靠性、完整性、可读性等等，较其它有联系的文件而言，元数据的相互关联，反映某项活动的外貌所要求的各个文件元数据的相互协同都需要控制。这就要求电子文件管理系统以及人为控制需要确保整个生命周期相互关联的各个文件的一致性问题。来自不同生成环境，不同电子文件管理系统的电子文件的元数据格式的互通性，统一性和同一性都构成了一致性问题的子问题。我们需要认清的挑战是，一致性因为时间和空间上的不同步难移保持，特别是在当下以用户体验为主的软件交互设计模式中，文件的生成很多是异步的，以此不会对用户体验产生阻塞，但这就造成了时间上的不一致性，使得相互关联的文件在同一时间点上处于生命周期的不同环节，这样对于保持相互的一致性无疑是一项极具挑战的任务。好比正常情况下，你要求一个大人和小孩跑得一样快，但是小孩却无能无力。还有相互关联的文件处于不同的流通环节，已收到文件和未收到文件的一致性也是较难保持一致的。

## 4．3电子文件信息的连续性问题及对策

从文件生命周期而言，电子文件从形成到运行的各个阶段都具备不同的状态，而这些状态并不是相互独立的，反之存在很强的前后关联性和整体连续性，对于确保各个节点的前后关联以及整体连续的同时为注入的档案属性也保持前后关联和整体连续也是十分有难度的，另外将归档内容完完全全前置也是不理想的，单单从这点对于就会让连续性受挫。

在《电子文件管理暂行办法》中也明确规定：电子文件传输、交换的时候，应当遵循相关要求，对传输、交换过程予以记录，这些记录将是归档内容的主要组成部分。电子文件传输流通过程中，大多数会经由不同的软件系统和硬件技术环境，那么不同的软硬件环境不可能做到从管理和技术方面的一致性，会经常发生一些动态的变化，显然每个环境生成的传输记录对应于不同的形式，面对不同的形式，前端控制能否保证和电子文件信息连续性相互映射的归档工作，也十分具有挑战。

根据电子文件的特点，电子文件在整个运行过程中所展现出的动态特性是十分明显的，对于具有多变、多态的特性的电子文件，将归档工作前置，并完完全全做到万无一失，整体连续，保证电子文件的价值和效用不会流失，如凭证价值需要大量的物力和人力耗费，对于领导的管理方式和工作者的耐性也是一场考验。

大数据时代的到来，电子文件生成速度之快，让电子文件的信息价值的时效性散失的也快，那么如果前端控制施加影响的时机不够及时，切入点不够准时，会导致文件生命周期各个阶段的信息搜集不够完备，元数据不够完整，将不能反映电子文件信息价值的连续性，不能涵盖必要细节。

总之，在实际工作中，后续归档工作和前期工作将出现信息难以及时衔接的问题，对应的各个部门的相互职责在短时间内也难以划分和有效对接，这对于保证电子文件信息连续和对应的前端控制内容连续都是要面临的问题。

## 4．4电子文件的知识化利用问题及对策

电子文件不仅具有长期保存上的优势，具体体现在持久化方式多样和居住容器的可观数据容量，最为重要的优势是较传统文件而言，更便于传递和利用。如通过检索技术能够快速获得目标信息，因而整体上能够促进知识获取和利用。加之一些研究方法，如文献计量和统计分析，大数据分析和挖掘，以及分析工具的使用，能够自动批量获取基于关键词的特定领域知识，模仿类似网络蜘蛛爬虫的模式，只不过攫取的对象是特定范围内或特定数据源中的文件。对于获取的文件内容，其质量高低和生产知识的潜能力，很大程度上取决于电子文件的分类和攫取算法的优越性。

将大量的电子文件进行有效的分类并注入档案工作色彩也是前端控制中需要实施的一项任务，一些指导分类工作的指导性原则，如《中国档案分类法》等相关的分类标准，虽然有助于电子文件的分类，但是鉴于历史主义和逻辑主义的不同应用场景和各自适应性，电子文件的分类工作不能被轻视和按照传统指导原则不加创新，随便进行分类，否则对于后期的保管、检索和利用将是巨大的隐患。数字化环境下，对于电子文件的分类，在一些基础原则的指导下，应当注入符合自身特色的创新内容，让分类后的电子文件也能够保持相关元数据内容的可回溯性和关联性，保证物理上和逻辑上的统一协调。而且对于相关软件应用也有好处，如网络爬虫攫取元数据内容，从而生成便于利用的辅助工具。基于这点，在后文的应用案例一篇，笔者借助火狐浏览器插件展示一个元数据构建良好，内容结构合理的文件将能带来怎样的好处和便捷，从而为知识利用打下良好的基础。

要确保电子文件的连续性和对应元数据在物理上和逻辑上的统一，必须利用恰当的技术措施在捕获元数据后，不断对文件进行集成，集成在这里是指融合、聚类和重组构成电子文件的要素，包括电子文件的内容信息、结构信息和背景信息，通过采用一定的技术标准和规范以及编码从逻辑上或物理上使其重新集成一个新的有机整体。

这里因为集成，将会对后续的电子文件聚类分析以及知识化利用产生影响。而且前端控制的介入使得电子文件的集成将遵循一定的档案原则，才能保障档案工作的前置。但集成后的电子文件将不便于聚类分析，因为聚类分析在文件处于分散的状态的时候是不便于获得整体的把握，但如果随意拆散原来的集成状态又会导致前端控制工作量的增大，导致前功尽弃。换言之就是前端控制的介入要求的文件的集成与后续利用对文件的聚合分析相互冲突。

## 4．5 电子文件前端控制的核心

对于电子文件的前端控制，核心是捕获其元数据并随着电子文件运动轨迹的变化，确保电子文件元数据的可靠性、真实性、完整性、可读性。而且很容易产生疏忽的地方是，每当我们提到元数据可能就只会想到电子文件的元数据内容就是记录其在软件环境中轨迹的记录，而遗失了其有档案色彩的元数据。这对元数据的完整性无疑是一种侵犯，对于元数据的完整性要求，也不单单是在数量上的齐全，还包括内容上的完整。所以，在记录电子文件各个运行阶段所对应的元数据的过程中，应当时刻明晰元数据的范围，防止元数据内容的缺失造成电子文件的自我描述性的降低，这对于软硬件元数据的缺失而言，有些较为优秀的电子文件管理系统还是能够在一定程度上进行文件元数据的追溯和恢复。但是对于基于档案色彩的元数据而言，机器终归是机器，没有人为灵活的去记录和收集，将对电子文件的元数据捕获工作造成重要问题甚至是前功尽弃。

在捕获元数据的同时，应该减少归档前置工作与文件生成、发布、流通以及利用环节的并发操作或基于一致性问题的异常捕获机制，再进行并行操作。即要么减少导致问题发生的不当操作，要么对可能发生的不当操作提供异常反馈和处理机制。对于由电子文件信息连续性带来的问题，应该借鉴通信领域抽样化模拟信号的方法，在万不得已的情况下，抓重要的忽略次要的，归档工作前置的时候，如果电子文件的动态特性如果表现的异常激烈，对应的档案前置工作不可能完全和电子文件运行轨迹一一映射的情况下，我们可以在一定的区间采用通用的处理方法去记录元数据，这样也是遵循成本和效益的平衡关系的。对于电子文件的知识化服务问题，我们应该在利用基础分类方法之上，应该考虑到后续知识利用需要聚合分析，从而在保持归档前置工作质量的同时也要防止文件集成过于耦合、过于紧密而不利于知识利用。

# 五、前端控制技术和方法

以电子文件特性和运行规律为基础、以前端控制理论为指导思想，对前端控制的技术和方法的探讨正是实现闭环的重要一步，同时为前端控制理论的应用实施积累宝贵的技术经验和方法总结。

## 5．1基于云计算实现批量文件前端控制

云计算可谓是当下的香饽饽，本质上它是互联网底和层基本设施的抽象，面向用户来说，它能提供基于互联网动态且能扩展的虚拟化资源，同时也是一种基于互联网的相关服务增加、使用和交付的一种模式。

它的超大规模的服务器集群，多样化的虚拟资源，高可靠性的镜像服务等等特点为应用的构建和迭代几乎提供了坚实的基础，最重要的是它拥有着惊人的计算能力。而正是因为这点，让档案学科依赖的技术解决方案的选择也对此积极拥抱。并不是因为档案学科一味追求新技术，而是着实基于自身考量，有积极拥抱的必要。

比如说档案资源因为档案馆馆际共享需要批量传递数字化副本，利用云服务的强大能力和低服务费用，将有利于控制档案馆成本和效益。云计算不仅为信息传输提供了超强的能力，还能够为档案馆间或各个部门之间的信息同步提供了重要保障。这对于前文探讨文件一致性问题的解决提供了技术上的突破口。各个部门可以建立和享受一块公有云，利用信息同步对分散的文件进行元数据等必要信息的上传、确保归档号的分配是建立在一系列相关文件形成的整体上，另外对于电子文件的知识化服务和利用问题，也可以借助公有云，将电子文件的的元数据等必要信息进行同步和上传，从而在云上做一些聚合分析，而无需在实际中，将已经在归档前置工作中安排妥当的大量文件从原来的集成状态变成聚合分析所需要的分散状态。因为对于知识服务利用问题，我们需要遵循的是逻辑主义。我们只需要对集成在各自对应的整体中的文件进行必要数据搜集，就可以对这些数据进行聚合分析以便实现知识服务和利用，而无需非要将实体聚合才能作分析。这无疑是一种物力和人力的浪费，造成了劳动力的冗余。

同时通过公有云天然的协同特性，很容易做到电子文件前端控制系统的分布式架构，从而实现电子文件前端控制的分布式处理和协同工作，另外在职责接口中明确界定分布架构中各个子部门的职责权限，让各个部门各司其职，各尽所责。一旦发生问题，比如说元数据缺失和不完整，能够快速追溯责任主体。这对于加强电子文件前端控制系统的反馈性将起到重要作用。在利用公有云的基础上，也需要时刻警醒的问题有：

确保电子文件前端控制系统在进行分布式处理的时候造成的数据不一致、同时操作带来的数据损坏等问题。因为当各个部门同时对电子文件的元数据进行上传和同步或者下载的时候，因为网络、软硬件设施等问题会使得电子文件的相关网络事务进行大量的并行操作，如果没有一个有效的控制电子文件事务先后队列或并行的机制，在利用技术带来的便捷性的同时一定会尝到技术双刃剑带来的严重苦果。

例如构成元数据整体的子元数据不属于同一文件生命周期阶段，这就会导致元数据整体损坏，那么一定需要设置校验机制，及时给各个部门提供反馈，并阻止其元数据同步和上传事务的提交，防止被整合到其它健康元数据当中，对其它元数据内容造成污染。虽然这种校验有利于从源头上杜绝问题产生，但是我们不能大意和疏忽，也应该对意外情况进行全盘考虑，实现双重安全机制。也就是在元数据同步事务当中，不同阶段的元数据发生整合后，一是需要我们实现一种恢复机制，也就是利用事务原子性，对不当事务进行撤销后能够恢复到原始健康状态。二是也应该意识到，元数据是具有连续性的，某个阶段元数据整合的不一致性不仅仅会影响到当下，还会对前面和后续的元数据整合造成不一致性问题。那么这个时候就需要我们建立元数据整合的日记系统，对于哪个阶段发生了不健康整合，我们就能通过日志查询发现受到该不健康整合事务影响的有哪些阶段，从而实现整体的健康恢复。

公有云和私有云的有效整合才能够处理好电子文件前端控制工作，而不是单单利用公有云的协同特性就可以将所有事务整合到公有云中，比如说网络星状拓扑结构，就只有一个中央节点和各个节点对接，基于这种结构的电子文件前端控制系统虽然具有控制简单、容易实现故障诊断和隔离、方便服务、网络易扩展等优点，但是中央节点的压力过于庞大，一旦出现问题，会导致整个系统的失灵和宕机。在电子文件前端控制系统的具体运作表现就是各个部门的子控制系统相对简单，职责相对单一，但不便于对元数据同步上传工作进行过滤和预处理操作，从而加大了系统整体上的不可控性。而且中央节点，也就是电子文件前端控制核心会集结大量工作，对于软硬件要求会很高，虽然有利于单一维护，但是职责过于集中，会导致各个部门失去积极性，任务简单而不利于从整体上提高前端控制工作质量。而且中央节点和各个节点的连接过于直接，将不利于重要数据的传输安全，特别是商业环境下的电子文件数据。解决的办法就是各个部门在共享公有云的前提下，建立与之对应的私有云，将能够大大减轻中央节点的工作量和确保系统整体上的健壮性。将元数据的校验工作分散在各个节点，而不是统一由中央节点集中接受和响应。同时对于元数据传递安全，将建立私有云和公有云的共同验证体系，私有云负责简单的校验，如元数据的完整性校验，这里可以封装下面将会探讨的完整性校验算法。当私有云通过校验时，再传递到公有云，对元数据的档案属性进行遍历和深度验证，按照重要性和必要性实现校验工作的分离。私有云对于确保各个部门的安全也是很重要的。

总的来说，对于云计算提供的能力，一部分是基于其为电子文件前端控制系统的架构来考虑的，特别是其协同特性和易于拓展的特性，另一部分是基于其基础设施之庞大，从而能够加速各个文件部门与核心管理部门之间的电子文件元数据内容、控制信息、反馈信息等构成前端控制信息体系的各种信息元素的传递和同步共享，也能同时实现批量文件的并行处理和操作。能够大大提高电子文件前端控制系统架构健壮性、功能可拓展性、数据安全性，同时也能够促进各个部门高效运转和电子文件前端控制的办事效率，也能从技术上、职责上清晰界定各个部门的职责，防止各个部门因为职责界定模糊导致相关工作无人处理，无人负责的尴尬局面。接下来会从电子文件前端控制细节来展开。

## 5．2利用MD5信息摘要算法确保文件完整性

MD5，即Message Digest Algorithm MD5。实质上，它是计算机领域实现符合预期定义的某种算法。形式上就是一个散列函数。它具有以下特点：

压缩性，任意长度的数据，经过散列函数算出的MD5值长度都是固定的。

便于计算，也就是从原数据经过散列函数算出MD5值得实现算法要求的计算量并不是很复杂。

抗修改性，对于原数据进行任何改动，哪怕只是一个字节，所得到的MD5值都有很大的区别。

强抗碰撞性，已知原数据经过散列后得到的MD5值，想要找到一个和该值相同MD5值的数据，即伪造的数据都是非常困难的。

可以将MD5看成给文件加上一个数字指纹，这个指纹的外貌都是差不多的，而且计算的简单性可以看成像用手指按一下一样易操作，而抗修改性通俗来说就是对于同一份电子文件，修改前后都由一个指纹差别非常大的人去分别按一下指纹。强抗碰撞性可以理解成对于附有指纹信息的电子文件，很难再重新拷贝成一份和其相同的电子文件，除非电子文件副本和原来文件在字节上都能够完全等同。即电子文件的不可伪造性。这点和拷贝电子文件是两回事。

在对MD5从定义和特点以及结合文件环境上认识之后，MD5对于确保电子文件的真实性、完整性以及相应元数据的可靠性、完整新都是极具保障的，当然不排除字典搜索等技术对抗算法破解的可能，但是对于现有环境下的电子文件管理要求，MD5的应用从技术上来说还是非常可靠的。

对于MD5的具体应用场景，包括电子文件数据传递方面和接下来的数字签名。从电子文件的特点来看，电子文件的信息和载体的可分离性，以及电子文件的产生环境情况各异。

对电子文件的完整性构成了极大的不确定性，包块内容上的不完整性和结构上的不完整性，而从这两个方面，在了解MD5的原理之后，显而易见的是都可以去解决，内容上的不完整性在物理层的体现就是缺少内容相关的字节串，结构上的不完整性可能表现为就是字节序列的顺序错误或某种结构内容的字节序列的缺失。总的来说都是字节的改变，那么按照MD5的定义，字节的任何一处改变都会导致散列函数映射出的MD5值产生巨大的差异，从而导致电子文件的完整性受到破坏。

在各个部门对电子文件进行前端控制的时候，例如归档工作前置时为电子文件注入档案属性的时候，或者按照档案工作要求对电子文件进行处理之前对原文件进行一次MD5散列映射：

即MD5(电子文件)，从而得到一个类似38b8c2c1093dd0fec383a9d9ac940515的Md5值。在注入档案属性的时候，如果发生错误需要回溯，可以在回溯之后再执行MD5（电子文件），再次得到一个类似38b8c2c1093dd0fec383a9d9ac940515的值。判断两次映射值是否相同，依据结果判断电子文件的回溯情况以及当回溯情况不同时，采取正确方法来知道回溯和恢复操作。再一利用场景就是，在基于上文中云系统架构的电子文件前端控制系统中，在各个节点向中央节点或者私有云对应的节点传送数据之前对电子文件进行一次MD5（传送数据）映射得到类似38b8c2c1093dd0fec383a9d9ac940515的十六进制串，在私有云节点中对传送过来的电子文件数据进行处理之前做一次校验，确保电子文件完整性未受到破坏之后再进行后续环节的其它处理操作。即MD5（传送数据）映射得到类似38b8c2c1093dd0fec383a9d9ac940515的十六进制串。如果发生不一致的情况，将阻止电子文件数据的进一步提交，并与此同时停止同一性质、同一时刻的其它前端控制操作，包括不同部门，不同物理设备上的相关应用程序，并利用信号机制快速上报信号处理系统。向遵循观察者-订阅者模式的各个相关系统及时发布异常信号，并由各个相关系统调用异常处理机制，在整个异常处理完毕时，再反馈异常处理完毕信号，最后对异常处理程序处理后的电子文件数据再做一次MD5映射，判断结果是否与原数据对应的MD5值相同，从而继续后面处理环节。

在对电子文件完整性鉴定应用MD5的时候，要注意一些应用操作的正误性对结果造成影响，虽然MD5原理是封装在电子文件管理系统当中的，但是应用系统不会对MD5原理只作一种功能，即不会硬编码一种用途，大多数是通过应用系统的配置或操作条件进行MD5的多种功能实现，例如对于一份电子文件，可能会包括数据内容、元数据内容、或附加数据等等，映射情况也各不相同，比如说你是MD5（数据内容）用于校验数据内容完整性、还是MD5（元数据内容）用于校验元数据内容完整性、还是MD5（附加数据等内容）用于校验附加数据内容的完整性。当然这些映射情况的选择都是根据电子文件管理系统的图形化界面来操作的。另一种形式是通过图形化配置界面，来配置MD5在应用系统中的初始配置选项，比如说是根据某一类文件应用哪一种映射情况，还是配置MD5某种应用配置的生命周期，映射容量，即映射多少份文件后检查映射配置是否和原先一致，来确定配置在时间上的互操作性，没有被改变，从而防止技术上的前后不一致性。确保电子文件在某个阶段所采用的校验手段是一致的。

另外就是某个子系统或者某个部门采用的MD5配置必须和其它系统或环境所采用的MD5配置是相同的，防止技术采用手段不同导致的结果误判，对于本符合校验完整性要求的电子文件数据内容，将其归结为错误、经过篡改的电子文件数据，造成不必要的循环检测，以至于不必要的人力、物力浪费。所以在启动某个阶段的电子文件数据传递的时候，一定得同步各个系统的技术参数，并做一些必要的单元测试和集成测试，以便于系统的各个接口数据类型、应用配置参数都是符合要求的。当然这和软件开发流程中定义的软件测试还是有区别的，前者更加注重业务上的测试，即电子文件数据传递业务的正确性和可用性、可靠性等。因为各个部门的业务状态是经常变化的，这种测试频率相对而言是较高的，测试的时间节点也是较为不固定的。而软件开发中的测试，只会在软件开发环节中，对于某种固定功能或者用户需求进行测试，这种测试只会发生在软件功能拓展和迭代中，相对而言测试的时间动态性是较低的，并且时间节点是较为固定的。弄清二者间的区别，为各部门的业务的正确执行打下良好的基础。

## 5．3采用新型数字签名算法确保文件真实性

在人类向知识化时代发展的过程中，数字资源的集成管理和集成服务显然成为一种趋势和必然。而电子文件在文件生命周期内也会不断的发生集成，具体包括以下几个方面，一是集成电子文件和相关的管理政策、法规、规划、评估机制，从而确保电子文件在不同处理阶段、不同行业领域、不同管理层级下的集成。二是，在软件开发设计中，将电子文件的管理要素融入软件设计当中，这不仅仅是开发业务功能，更需要了解管理环节，充分利用已有经验，融入特定的管理要素，以至于能够反映特定环境下的业务背景。三是，在所有电子文件集成服务的基础上，将电子文件与其它数据资源集成起来用于提供更加完备、更加系统、更加整体的服务。

那么在大量的集成操作中，如何确保各个集成元素的真实性，或者具有自我描述性，并且不仅仅能再现文件内容上的描述信息，更能反映电子文件的形成主体或者其它重要信息。应用数字签名成为一种重要手段和工具。

实现数字签名的算法和与原理多种多样，但是也不是什么数字签名算法都可以满足业务需要的，特别是电子文件等数字资源在时间和空间上的多元多维集成，对于数字签名算法的稳定性要求是比较苛刻的。

数字签名意如其名，类似普通的文档签名，只不过它是通过数字形式形成于特定环境、和电子文件相伴相随的一种电子文件附加数据。相比较于元数据，它可能更加偏向用于判断电子文件来源上的真实性，更具有法律效力。

数字签名在原有数据基础上通过附加一些数据、或对原有数据做密码转换形成的附加数据，在技术上，它需要借用类似上文中的MD5来确保数据的完整性，并且在此基础上利用加密技术来进一步实现数字签名的应用。因为应用数字签名不仅仅能够确保数据完整性、还能够确保电子文件来源的正确性、真实性。数字签名包括数字签名和数字签名验证两个环节，按照需要可以基于对称加密机制或者非对称加密机制。而后者更为适用于苛刻条件。

在电子文件前端控制中，数字签名的应用场景主要有确保电子文件来源和数据完整（因为其基于类似MD5算法的实现），为电子文件等数字资源的集成来源提供了可靠的保障。有利于确保电子文件形成主体的权威性和保护电子文件形成者的相关权益，以及当意外情况发生时，便于追溯和明确相关文件的责任主体。

例如在集成系统中，需要将电子文件和其他数字资源集成以提供更完备的服务、如前文中知识服务。就需要在集成的时候，对于参与集成或构成集成的各种电子文件以及其它形式的数字资源进行数字签名和数字签名验证，当某个数字档案馆准备对企业提供历史某个时刻有关其生产经营的经济数据时，在统计经济数据时，对于某份电子文件，在知道当时该企业的经营主体的或该电子文件的形成主体的情况下，就可以对这份电子文件进行数字签名验证，如果得到的数字签名信息与形成者情况不符，那么该份文件以及后续基于该文件的利用成果都将是不可靠的，并且需要执行后续处理环节。防止后续的电子文件信息来源重蹈覆辙。

数字签名虽然在电子文件前端控制系统中封装成仅仅借助于图形界面就可以实现的图形化操作。但是掌握其具体细节将有利于掌握前端控制系统的配置和熟练业务的操作，以及当意外发生的时候，能够快速排错并实现自我恢复。

就拿基于云系统架构的电子文件前端控制系统来说，当一个部门向中央节点或者私有云对应的节点进行重要数据的传递，并且有电子文件来源上的验证需求，首先该部门需要根据这份电子文件的原数据通过哈希函数计算出有关该数据的摘要信息，（基于非对称加密算法）

需要将计算出的摘要信息通过私钥进行加密，加密后的摘要和原数据一起进行传输，在传输到中央节点，也就是前端控制系统核心或者私有云对应的节点，也就是前置校验节点，在处理和利用之前，需要利用和生成数据摘要同样的算法对传送来的数据进行摘要计算，再利用和私钥相对的公钥对传送过来的摘要数据进行解密，生成的摘要和前者进行对比从而判定数据来源是否准确，也就是电子文件的主体是否是正确的。

对于数字签名在电子文件前端控制系统的应用，需要我们注意的问题有：

一是和MD5应用系统配置一样，确保各个系统或者各个节点有关数字签名的配置是完全相同的，防止不一致产生的错误验证结果造成电子文件前端控制的失误。

二是充分了解数字签名图形化操作甚至是其背后运行原理的基础上，确保数字签名的时序问题，防止因为技术操作时序不合理对数字签名验证结果造成误判。

三是防止数字签名和数字签名验证后的摘要信息泄露给前端控制系统埋下安全隐患以及造成个人信息泄露。

四是根据管理需要和系统安全适时为电子文件前端控制系统应用新型签名算法，这就要求电子文件前端系统的设计中，应当防止数字签名模块和其它模块过于强耦合。MD5实现模块亦是如此。

## 5．4为电子文件添加语义性内容增强可读性

对于电子文件的自我描述性不仅仅要求其具备特定的内容数据，还需要其自我描述内容具有能够被应用工具识别的结构，从而展现出更加便于利用的可读性。因为电子文件的数量是可观的，而不是单份的，其处理也不是逐件的。正是鉴于其可观的数量以及批量处理的方式，更需要借助应用工具对其语义进行分析并总结以便于知识化服务或利用。

同时认知时代的快速发展，人们对于知识消费表现出急切的渴望，但是这种为知识买账的交易是建立在质量上乘的知识内容基础之上的，所以对于应用于知识化服务和利用的电子文件来说，前端控制为其不断注入可读性，也就是在精选质量上乘的原材料，确保知识化利用成果和同类产品比较起来更具有竞争优势和更好的服务利用。

对于电子文件的可读性，可以从电子文件的构成要素，如数据内容、元数据信息、结构内容以及背景信息等不同方面着手，注入各自对应的语义性内容。对于特定的要素需要采用特定的格式去注入增强可读性的内容。并确保对各要素的可读性内容的格式做好格式说明，而且对于格式的采用，应该尽量选择通用性较强、应用工具能够识别的格式，而不能够为了标新立异创建一种独具一格的格式，同时需要考虑到前后兼容，也就是旧系统和老系统都能够统一里处理和利用，或者通过适配器模式做出简单的转换之后也是能够可利用的。在后面的实践中，就采用了基于数据内容的可读性增强操作，以便于特定应用程序能够捕获并生成特定的可读信息。

在保证可读性注入内容的格式通用性、兼容性、易处理性基础之上，对于注入的内容也要符合前端控制要求，因为前端控制工作的连续性，也要确保注入内容的连续性，以便在整体上构成一个可生产的可读性内容整体，而不能因为某部分注入内容的缺失导致可读性生产中断以及可读性不完整，相对而言显得不那么可读，也就是可读性内容的不可读。

在保证可读性注入内容的连续性的基础上，应当对注入内容的内容本身做出一些精选甄别，以保证生产出的可读性内容能够反映出电子文件的一些情况，便于电子文件可读性应用工具的处理以及最终的知识化服务和利用。前端控制中，在为电子文件注入可读性内容的时候，需要注意的问题有：

一是确保技术手段的前后一致性，即使不一致，也至少保证不同的技术手段产生的可读性输出是一致的，这样有利于减轻处理问题。

二是可读性内容的注入不能一时兴起，也需要建立相关标准确保不同阶段、不同时间节点自动或手动按照特定格式生成。

三是可读性内容的生成过程不能延时。在于电子文件的前端控制系统中，不同的处理节点应该在上游节点的可读性内容基础上，进行新内容的注入的同时为下游节点的语义性添加工作做好必要的基础，如特定的格式说明和技术参数说明

# 六、有关前端控制系统的建议

前端控制技术和方法的研究成果和利用很可能会被集成和封装进电子文件前端控制系统，对于其往往需要忽略技术细节、只需要管掌握图形化操作利用，更需要遵循前端控制系统的使用原理和规范，才能真正发挥前端控制系统的作用和防止出现意外。

## 6．1电子文件前端控制系统技术上的建议

从电子文件数量的增长趋势来看，电子文件的处理方式的转变势在必行。而前端控制对于确保电子文件的真实性、完整性、可读性，以及提高归档电子文件的质量起到了重要作用。所以，前端控制的方式也应当随着电子文件的处理方式的转变而转变，随着计算机水平的高速发展，软硬件水平不断提高。前端控制系统建设已经成为前端控制的重要保障。一个良好的前端控制系统，将为电子文件的前端控制工作的顺利开展提供良好的开端和保障。基于技术的视角，给前端控制系统的建设提供以下建议作为参考：

一、仔细辨别电子文件管理系统和前端控制系统之间的联系

电子文件管理系统为前端控制系统提供了文件生命周期内文件的集中管理。而电子文件前端控制系统是有别于此的，在于对文件施加管理的过程中，更加注重归档前置工作内容的注入。所以从技术上，在电子文件管理系统的基础上，电子文件前端控制系统应该涉及更多的模块内容和更加完善的字段体系，如归档号的分配。更加强调电子文件的档案属性。

二、电子文件前端控制系统在设计时应更加强调维护性、可重用性、模块性

因为基于云架构的电子文件前端控制系统架构的复杂性，在软件设计之初时，应当摒弃传统事业部门的设计特点，很多事业部门因为经费不足、领导的不重视，对于一些管理系统的设计非常的草率，结构简单，技术过时，只注重功能实现，并不重视功能的可拓展性以及技术的先进性，实质上，对于这种现象，虽然情有可原，但是从长远来说，并不利于长期成本和效益的控制，甚至当管理政策、规范变动时，新管理系统的开发成本将大大的超过一个起初就设计精良的系统。

因此对于电子文件的前端控制系统的建设，应该更加注重维护性、可重用性、模块性，鉴于电子文件的管理规范、技术参数变动更为频繁，电子文件前端控制系统不能用传统的管理视野去一视同仁。应当在经费允许的情况下，积极追求系统的优越性。

三、电子文件前端控制系统应当注重迭代更新创新

在基于模块化设计的基础之上，应当根据需要，比如说类似MD5算法的新型算法的替换工作，当比MD5算法更为优越的算法用于确保电子文件完整性出现时，或者更加优越的新型数字签名算法出现时，应当对原有电子文件前端控制系统做出迭代，但需要注意的是，并不是对原有系统整个替换，而是根据模块化分析，替换对应模块。这样将有利于成本效益控制。在模块化分析的时候，也应当注重模块之间的耦合度分析，防止模块替换导致模块接口不一，从而导致电子文件前端控制系统的崩溃。

## 6．2电子文件前端控制系统管理上的建议

一个良好的电子文件前端控制系统能够为电子文件前端控制工作提供坚实的基础，虽然电子文件前端控制系统在设计之时已经内置了一些管理要素，比如说权限控制、身份校验、流程控制等等。但是这些毕竟是硬编码进去的，对于前端控制工作中出现的各种各样的情况的复杂性而言，这些硬编码的管理要素是不足以应对的。基于管理视角，为电子文件前端控制系统提出以下几点建议：

一、制定电子文件前端控制系统的管理规范、确保各部门管理口径一致。

为电子文件前端控制系统制定相关的管理规范，如技术参数规范、操作人员权限配置管理。这种管理规范不仅仅是面向部门内部的，而且需要同步到整个系统。鉴于各个部门的管理规习惯的不同，如果不实现管理规范的同步，将不利于各个系统的数据同步、更加不利于基于云架构的电子文件前端控制系统的各个节点的无障碍交流。加强管理规范的建设和同步，将有效减少因为管理规范的差异过大而带来的各个部门系统交流的不一致性问题。

二、防止电子文件前端控制系统管理权限过于集中

电子文件前端控制系统包含重要的安全成分，如数字签名库含有重要的个人信息，以及其他重要信息。如果将电子文件前端控制系统的管理权限过于集中到一个人或一个部门的身上，不但不利于全部门人员的系统学习，也不利于信息安全。但与此同时，管理权限也不能过于分散。软件管理不同于传统管理，软件中日常的数据流动是比较频繁的，因而重要价值信息所面临的风险也是较大的，过多的人拥有管理权限将导致风险增大。同时也不利于管理事务的集中审核审批，需要耗费较大的人力才能处理好某一项管理事务。

三、加强各部门有关电子文件前端控制系统管理经验的分享交流

各部门有关电子文件前端控制系统的经验交流和分享不但有助于管理规范的统一化步伐，而且也有助于借用头脑风暴，为系统的改进提出创新见解，能够为推动电子文件前端控制系统建设与前端控制理论相结合起到重要作用。

# 七、前端控制应用案例

要想深入贯彻前端控制思想在前端控制工作中的应用，关键在于实践。不断的实践能够为前端控制运用提供大量的经验反馈，以此作为风向标不断对理论进行修正，再利用修正结果指导实际前端控制工作，久而久之，将有利于形成理论和实践的互利互惠的局面。接下来将从三个实例展开前端控制研究。对于涉及到的技术原理会做简要说明，虽然没有深入细节，但旨在通过视觉呈现为前端控制提供技术上的可行性验证。

## 7．1 HTML5 Canvas API图片处理与前端控制

HTML5 Canvas API 主要是为Web端的图形化绘制提供了编程接口，从而可以利用脚本语言，通俗来讲，就是我们可以做到绘图、图片合成、简单动画等功能。确保电子文件的完整性是前端控制目标之一，结合Canvas的图片相关功能以及前文的MD5算法，让我们来验证一下MD5算法对于保障电子文件完整性的作用，通过视觉加深MD5算法的在电子文件前端控制中应用的可行性。

主要实践操作就是，先准备一张图片文件来模拟一份电子文件，然后运用MD5对其进行映射，即MD5（图片文件）得到一串十六进制值s1，然后对该张图片施加几类操作，用于模拟对电子文件的篡改行为，然后对施加几类操作后的电子文件分别进行MD5映射，分别得到s2、s3等。比较s1和s2、s1和s3是否相等，是否符合MD5验证完整性的预期行为。从而验证MD5算法在电子文件前端控制的应用可行性。

其中为了视觉分析，特意利用放大研究方法，将细小的篡改放大，以便于分析和得出结论。在部分修改操作中修改了比较多的字节。但需要知道的是，仅仅修改一个字节的图片信息是可以的。

## 7．2 Web Audio API音乐可视化与前端控制

Web Audio API 主要是为Web端的音频操作提供了编程接口，从而可以利用脚本语言，通俗来讲，就是我们可以做到音乐可视化，音乐播放，音频文件操作，音频特效制作等功能。确保电子文件的完整性是前端控制目标之一，但是对于一些自我描述不会破坏文件的已存在应用或者系统，我们不能完全信任。仍然需要对其功能进行验证，以保证被其处理的电子文件完整性不会受到破坏。结合Audio的内置analyser节点，它将自己描述成只会分析音频文件的数据做可视化，但是不会坡改音频文件的数据。就此描述，结合MD5算法我们来做一次面对音频文件的完整性验证。从而提高了MD5算法确保完整性的应用普适性（不但可以保障图片文件、也可以保障音频文件的完整性）

其中，我们将音频文件模拟电子文件，将analyser节点模拟自我描述不会破坏数据的应用，接着使得音频文件数据流向该节点，模拟电子文件被某个应用处理，在流向之前做一次MD5映射，即MD5（音频文件），得到s1，接着在流向之后，再做一次MD5映射，得到s2，比较s1和s2.从而判断该节点（模拟应用）是否具备和自我描述的保护文件完整性的行为是一致的。

注意我们通过可视化展现电子文件内部数据的形式，并通过视觉分析来得出结论，对于技术深层细节做必要忽略，但不会影响验证。

## 7．3 通过火狐浏览器查看文件语义化带来的可读性

本文探讨的语义化有两种含义：一是指内容具有可读性，二是指电子文件具有特殊的标识或者按照某种算法使得应用工具、程序能够识别并生成可读性的内容，例如大纲。

该实例结合HTML5 Outline大纲算法，对比两份HTML文件（一种电子文件类型，即网页文件），其中一份文件应用大纲算法，而另一份没有应用，并同时用火狐浏览器插件outliner去分析这两份文件，从视觉上带来语义性电子文件给利用带来的便捷性和设备可访问性。从而为电子文件前端控制在保持和增强可读性方面提供一种借鉴。

可以通过多种形式和手段来进行电子文件的语义化，本实例用<h1>、<h2>、

<h3>等html标签结合大纲算法来进行语义化和语义化利用，并从视觉上展现可读性增强效果。

# 致谢

对于很多同学来说，四年的大学生活即将划上一个句号，但在图书情报档案系的风雨洗礼后，让我发现我的大学生活并不能用一个句号去结束，我想用一串省略号代表仍炽热在心中无限感恩和遐想。

初来乍到，我们像嗷嗷待哺的孩子，对于一切都是充满着新奇，学术的匮乏让我们充满了求知的欲望，而且各位老师并没有让我们感到失望，在我们面前是一群学术造诣颇深的园丁、更重要的是他们仍然迸发出年轻、青春的活力，让每一节课堂都激荡着澎湃的浪花，尽情滋润着稚嫩的我们。

大学四年，我从一个渺茫的懵懂青年，在各位老师的做人、做事思想的熏陶下，并在刻苦不懈的努力下，逐渐抓住了青春的尾巴，在IBM找到了自己的一席之地。虽然最终没有从事档案的相关工作，但时时刻刻都感受到了图书情报档案系给我带来的有关为人处世、生活学习的巨大收益。

写到这里，我想借用一首诗歌去感恩最美的老师：我原想收获一缕清风，您却给了我整个春天。我原想捧起一簇浪花，您却给了我整个海洋。我原想撷取一枚绿叶，您却给了我整个森林。我原想亲吻一朵雪花，您却给了我整个银色的世界。您在最高点才绽放美丽，也把最美丽的时刻全给了学生，一种损失让您因为那刻骨铭心的美而获得了永恒。那绚丽无边的晓晨和一个渐渐远去的背影，已定格为我心中永远的一帧老照片！

最后，再次对各位老师的尊尊教诲表示感激之情，也对导师在论文期间，放下自身繁忙的事务，为我和同学不厌其烦指点论文表达由衷的感谢。愿各位老师奉献教育事业，几载归来之后，仍有一颗少年之心。

另外，金无足赤，人无完人。由于我的学术水平有限，论文之中难免有不足之处，恳请各位老师的批评和指正。

# 参考文献

[1]金鑫.电子文件呼唤前端控制[J].档案学研究,2004

[2]杨安莲.试论电子文件的本质属性[D].浙江档案,2013

[3]石玉.李肖军.电子文件前端控制理论浅析[D]华北电力大学,河北大学,2007

[4]赵屹.基于前端控制思想的电子文件形成过程研究[J].档案学研究,2012(3):16-23

[5]冯惠玲.电子文件与纸质文件的管理互动[J].中国档案,2003(12)

[6]许爱香.电子文件前端控制原则与机关档案室工作[J].山西档案,2003(4)

[7]金波.丁华东.《电子文件管理学》[M].上海.上海大学出版社,2015

[8]徐维.管志宇.电子文件前端管理的来源及其他[J].档案与建设, 2001,(03)

[9]吕元智.关于文件前端控制的几个问题[J].图书情报知识, 2002,(06)

# 附录

1 Canvas 应用案例截图

初始状态：



不作处理：



前后MD5映射值对比：



全部处理：

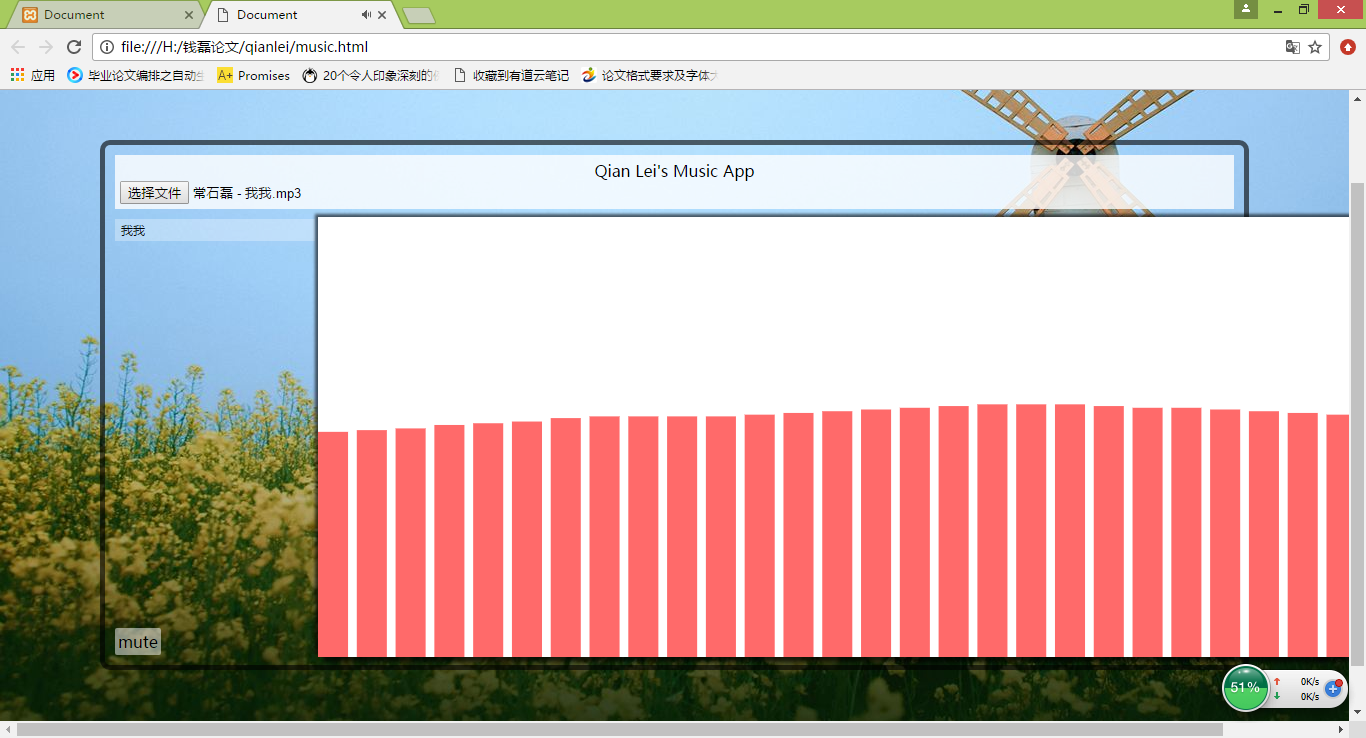


前后MD5映射值相比：



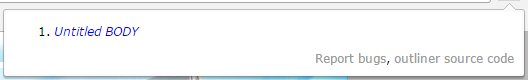
2. Audio 应用案例截图

音乐可视化以及文件完整性分析（请参照应用案例说明）：



3．H5 Outliner 应用案例截图：

上海大学图书情报档案系主页源HTML文件，缺少可读性注入，所以可读性很差





上海大学百度百科加强可读性内容注入之后，可读性明显，能被应用工具识别并自动生成大纲。

