

# 电子文件元数据分类与方案设计对策研究 课题研究报告

项目负责人：陶水龙

承担单位：北京市档案局

完成时间：2015年11月

国家档案局  
WWW.SAAC.GOV.CN

# 目录

一、研究综述.....	1
(一) 课题研究的意义.....	1
(二) 课题研究的内容.....	1
(三) 课题研究的对象.....	2
二、国内外电子文件和电子档案元数据研究文献综述.....	3
(一) 文献研究综述.....	3
1、电子文件元数据研究的基本情况.....	3
2、当前研究中对电子文件元数据类型的划分.....	14
3、电子文件元数据的作用研究.....	15
(二) 我国部分省市档案馆对于电子文件和电子档案元数据的研究、应用情况.....	16
1、制定标准接近统一.....	16
2、设计电子文件全过程元数据管理.....	17
3、元数据元素数量较多.....	17
4、元数据与著录项数据关系交叉.....	17
三、电子文件和电子档案元数据研究在实际应用中存在的问题与分析.....	18
(一) 电子档案接收在北京市的试点情况及其分析.....	18
1、试点情况.....	18
2、试点中发现的问题.....	19
(二) 元数据理论研究在实践过程中出现的问题.....	20
1、数字档案馆建设过程中, 电子档案元数据的分类标准认识不统一.....	21
2、电子档案的元数据集合的无序增长.....	21
3、电子档案元数据格式呈现多元化、复杂化的趋势.....	22
(三) 破解存在问题的理论依据.....	22
1、元数据是保证电子文件具有凭证作用的数据.....	22
2、文件连续体理论.....	24
3、“前端控制”思想.....	25
四、电子文件和电子档案元数据分开管理的设想.....	26
(一) 分开管理的目的和原则.....	26
1、明确使用电子文件和电子档案元数据的目的.....	26
2、明确保证电子文件和电子档案真实性元数据的本质特征.....	27
(二) 电子文件和电子档案元数据划分为对象元数据和事件元数据.....	28
(三) 概念界定.....	31
1、对象元数据.....	31
2、事件元数据.....	31
五、电子文件和电子档案元数据分开管理系统框架.....	32
(一) 电子文件和电子档案元数据分开管理系统框架的特点.....	33
1、立足档案管理需求, 划分档案元数据分类并进行分别管理.....	34
2、提供更加有效的元数据集合.....	34
3、支持实现元数据集合的质量控制.....	35
(二) 分开管理元数据的分类标准.....	36
1、实现元数据管理方便、操作便捷的目的.....	36
2、元数据划分类型逻辑性强并与著录信息区分.....	36

国家档案局  
www.saac.gov.cn

3、让不同使用对象认识更明确、操作更简单.....	37
(三) 分开管理的具体方案.....	37
1、方案项目设置的具体分析.....	54
2、文本类对象元数据具体选项的分析说明.....	55
3、文本类事件元数据具体选项的分析说明.....	59
4、图像类对象元数据具体选项的分析说明.....	62
5、图像类事件元数据具体选项的分析说明.....	66
6、影像类对象元数据具体选项的分析说明.....	68
7、影像类事件元数据具体选项的分析说明.....	69
8、声音类对象元数据具体选项的分析说明.....	71
9、声音类事件元数据具体选项的分析说明.....	72
10、元数据不同使用对象需要的不同内容.....	74
六、电子文件和电子档案元数据分类管理的应用前景.....	74
七、课题的重要研究成果.....	77
论文-电子档案元数据分开管理研究（发表于《中国档案》2014年第9期）.....	77
参考文献.....	81
表 1 试点单位移交电子档案情况.....	18
表 2 元数据类型.....	28
表 3 文本类电子文件和电子档案对象元数据.....	38
表 4 文本类电子文件和电子档案事件元数据.....	40
表 5 图像类电子文件和电子档案对象元数据.....	43
表 6 图像类电子文件和电子档案事件元数据.....	45
表 7 影像类电子文件和电子档案对象元数据.....	47
表 8 影像类电子文件和电子档案事件元数据.....	49
表 9 声音类电子文件和电子档案对象元数据.....	51
表 10 声音类电子文件和电子档案事件元数据.....	52
图 1 文件连续体模型.....	25
图 2 元数据管理框架的设计流程.....	33

## 一、研究综述

### （一）课题研究的意义

对电子文件元数据分类的研究是电子文件真实性保证、长期保存、科学管理、有效利用的基础，也是数字档案馆建设中的重要组成部分，也是北京数字档案馆（北京电子文件中心）工程建设、实施和广泛应用的基础性工作之一。从概念清晰、分类合理、使用便捷、科学保管、有效利用的角度进行分类的科学划分，是电子文件元数据研究的基础，也是其他涉及电子文件和电子档案研究的基础保障。

本课题旨在对电子文件和电子档案元数据，尤其是对与电子文件和电子档案凭证性、真实性等相关元数据信息和与管理流程相关的管理元数据信息进行重新解读，以厘清电子文件和电子档案元数据的相关概念与特征，提出电子档案元数据分开管理的理念，并研究建立适合电子档案元数据管理系统框架。

### （二）课题研究的内容

本课题通过调研国内外电子文件元数据的制定、研究和使用情况，分析研究并提出科学的电子文件元数据分类方法，并对分类进行方案设计，为确保电子文件的真实性、凭证性、长期保存、有效利用和科学管理提供基础支撑。

主要研究内容包括：

必要性研究。确定电子文件元数据分类的意义、原则、方法、标

准等。

现状调查研究。对国内外关于电子文件元数据的研究、标准、使用等方面内容进行归纳分类，提出本课题研究确定的参考建议。

比较分析研究。对国内外电子文件元数据分类从档案管理学、电子文件长期保存等角度进行分析、研究，探讨其对本课题的借鉴与参考，重点分析已有成果的不足与可优化的方面，形成本课题所提出的内容。

方法与方案研究。以精确、统一、便于利用等为标准，提出电子文件元数据分类方法，同时以突出电子文件真实性、凭证性，以强化电子文件长期保存、科学管理为目标，提出电子文件元数据分类的具体实施方案。

### （三）课题研究的对象

本课题以电子文件和电子档案的元数据为研究对象。

档案部门和相关信息主管部门关于定义电子文件和电子档案的问题，目前在国家层面和一些省市所指的并不完全一致，有时是根据需要随时转换。作为档案部门和档案工作人员对这两个概念的解释也存在多种理解。

关于电子文件和电子档案，北京市档案局定义并明晰了这两个概念的时间、动作边界。因此，在本课题研究中，遵循北京市档案局的相关界定，即：电子文件指的是在档案室归档前的状态，而归档后以及移交到各级各类档案馆中的统称为电子档案。

## 二、国内外电子文件和电子档案元数据研究文献综述

### （一）文献研究综述

选取国内外部分涉及电子文件元数据内容的标准，中国知网相关文章，美国和澳大利亚、国际档案理事会 ICA 涉及电子文件元数据方面的定义分类及相关研究进行文献分析。国际标准化组织、国际档案理事会（ICA）、美国、澳大利亚涉及电子文件元数据方面的标准规范，我国部分涉及电子文件元数据内容的标准，中国期刊网相关文章，对电子文件元数据定义分类及相关研究进行文献分析。

具体包括：ISO23081《信息与文献-文件管理过程-文件元数据-第1部分：原则》，以及第2、第3部分；ISO15489《信息与文献-文件管理-第1部分：通则》；国际档案理事会颁布《电子办公环境中的管理原则与功能需求》；澳大利亚政府文件保管元数据执行方针；美国；国家档案行业标准 DA/T46《文书类电子文件元数据方案》；《电子文件元数据标准》等的系列研究成果。

#### 1、电子文件元数据研究的基本情况

国内外有关元数据标准的研究发展比较迅速，国内相关标准包括国家标准、国家计划标准、行业标准、工程标准，国外相关标准包括国际标准化组织（ISO）、美国、英国、德国、欧洲、日本、法国、加拿大、澳大利亚、新西兰等国家行业元数据标准。课题组研究过程中参考的元数据标准据初步统计，了解到相关元数据标准共 212 项，其

中，国内元数据标准 58 项，国外元数据标准 154 项；有关元数据内容的标准 186 项，有关元数据格式的标准 26 项。

( 1 ) ISO15489 Information and documentation — Records management。ISO/TC 46/SC11 于 2001 年发布了 ISO 15489（信息和文献记录管理）系列标准：ISO 15489-1：2001 Information and documentation — Records management — Part1：General 和 ISO/TR 15489-2：2001 Information and documentation—Records management —Part 2: Guidelines。其中提出：文件管理元数据（records management metadata）是“用来描述文件背景、内容和结构及其整个管理过程的数据。<sup>1</sup>

( 2 ) ISO 23081 Information and documentation — Records management processes—Metadata for records

ISO 23081 Information and documentation — Records management processes—Metadata for records 同样由 ISO/TC 46/SC11 提出，包括 3 个系列标准，即：

ISO 23081 — 1：2006 Information and documentation — Records management processes—Metadata for records—Part 1: Principles（第 1 部分：原则）。《ISO 23081 第 1 部分，文件管理流程——文件元数据：原则》，于 2006 年 1 月正式发布，主要内容包括：文件管理元数据的目的和定义；作用和责任；与其他元数据域的关系；元数据管理。ISO 15489 要求的元数据类型。

<sup>1</sup>ISO 15489-1 2001, Information and documentation—Records management—Part 1: General

ISO 23081—2： 2009 Information and documentation—Managing metadata for records—Part 2: Conceptual and implementation issues（第 2 部分：概念及实施）。《ISO 23081 第 2 部分，文件元数据：概念及实施》，主要内容包括：业务背景；策略问题和责任；建立元数据框架；概念问题（例如，元数据集，元数据继承）；基本元数据元素；元数据方案设计和建构；实施问题如系统设计问题（例如：捕获，存储，保存）、元数据管理；独立实施检查清单。

ISO/TR 23081 — 3： 2011 Information and documentation — Managing metadata for records—Part 3: Self-assessment method（第 3 部分：自评估方法）。《ISO 23081 第 3 部分，自评估方法》，由三部分构成：元数据方案分析工作单、元数据方案分析工作单使用指南和元数据方案评估问题分析流程，依据 ISO 15489 和 ISO 23081—1 对现有元数据集进行评价，吸取了 InterPARES 项目元数据方案登记的研究成果。

ISO23801 文件管理元数据标准系列具有支持 ISO15489 文件管理元数据实施的特点，通过规范元数据管理原则、方法和评估工具落实 ISO 15489 文件管理的原则和目标，是 ISO15489 文件管理标准的进一步扩展，可以帮助人们从文件管理的角度理解元数据。<sup>1</sup>

### （3）都柏林核心元数据

1995 年 3 月，由联机计算机图书馆中心（OCLC）与国家超级计算机应用中心（NCSA）联合发起，52 位来自图书馆界、计算机网络界

<sup>1</sup>安小米等，基于 ISO 15489 的文件档案管理核心标准及相关规范[M]. 北京：中国标准出版社，2008. 2: 1:1-139

专家共同研究产生。目的是希望建立一套描述网络电子文献的方法，以便网上信息检索。后来形成 DC 元数据标准(都柏林核心 Dublin Core 元数据)，其基本方案是包括 15 个“核心元素”的集合，由都柏林核心元数据倡议（DCMI）负责维护。提出了内容描述、知识产权和外形描述 3 个层次 15 个元数据元素：

内容描述：关于资源内容相关信息的描述，包括题名、主题或关键词、资源描述、来源、语种、相关资源和时空范围。

知识产权：关于资源知识产权相关信息的描述，包括创建者、出版者、其他责任者和权限管理。

外形描述：关于资源外形特征信息的描述，包括日期、资源类型、资源形式和资源标识。

每一个都柏林核心元素都是可选和可以重复的。都柏林核心元数据启动计划（DCMI）已经建立了提炼元素和鼓励使用编码和词汇结构的标准方法。<sup>1</sup>

都柏林核心元数据标准最初的目标是描述爆炸式增长的网络信息资源，然而与其初衷偏离的是，该标准并没有很好的解决互联网信息资源的描述问题，相反，在描述一般信息资源方面，成为了很多应用领域制定本领域元数据标准的参考。都柏林核心元数据的内容需要结合具体应用进行大量的扩展，它的优点是灵活，缺点是很难保证系列标准的一致性。<sup>2</sup>

---

<sup>1</sup><http://zh.wikipedia.org/wiki/都柏林核心>

<sup>2</sup>宦茂盛，空间数据标准化的方法与应用研究，[A]. 北京：中国科学院遥感应用研究所，2010.

国家档案局  
www.saac.gov.cn

#### (4) 永久保护真实电子文件国际研究项目 InterPARES

国外关于元数据的研究课题较多，进展快，成效显著，最著名的就是国际档案界的跨国研究：“永久保护真实电子文件国际研究项目（InterPARES）”，包括中国、加拿大、美国、英国、爱尔兰、瑞典、荷兰、法国、意大利、葡萄牙、澳大利亚等 10 余个国家的档案馆及一些大学和研究机构共同参与研究。

InterPARES 主要是为了永久保护电子生成的文件及其真实性建立一套基本的理论知识和方法，为制订有关长期保管电子文件真实性方面的方针和标准规范提供指南。这个项目成立于 1999 年，有十余个国家参与，项目研究的领域主要分为四个部分：保护电子文件的真实性有何要求；怎样挑选应保存的电子文件；如何保管真实的电子文件；哪些方针、策略及标准可用来长期保护电子文件的真实性。InterPARES 提出的元数据模板包括五个模块：载体、文件类型的外部元素、文件类型的内部元素、标注、背景环境。

载体：载体的识别（现行载体、归档载体）、载体特征（类型、构成的物理材料、规格）、载体为接收信息所作的准备、载体的存取类型、载体的存贮密度与容量等。

文件类型的外部元素：人类的惯用语言、显现特征（又名脚本原件），包括总体显现特征和具体显现特征，总体显现特征如文本、图形、图像、声音、多媒体等、专门符号、印记（如受托第三方的文件法律效力的认证、由受托第三方加盖的数字化时戳、数字化署名等）。

文件类型的内部元素：责任者的名称、来源的名称、按年月顺序

排列的日期、文件形成的所在地的地名、收件者的姓名、被抄送者的姓名、活动（事务）的表达、活动（事务）的描述、执笔者的姓名、独立证明、署名的资格证明等。

标注：办文过程中所作的附注（文件传递的优先次序、文件传递的日期时间和地点、附件情况的说明等）、在处理与文件有关事务的过程中所作的附注（文件接收的日期与时间、处理办公机构的名称、需进行的活动或文件传递的日期与时间等）、为实现文件管理的目的而在文件管理中所作的附注（如归档日期、稿本或版本编号、成套或一类文件中的文件序号、成套文件编号、分类号、收发文登记号、文件形成者的姓名等）。

背景环境：法律行政管理方面的背景环境、来源方面的背景环境、活动过程方面的背景环境、文件方面的背景环境、技术方面的背景环境（包括计算机硬件，如主存贮器、辅助存贮器、第三存贮器、中央处理器、网络、外围设备、结构体系等；计算机软件，如操作系统、系统软件、网络软件、应用软件等；数据，如文件结构、数据格式或文件格式等，系统模型和系统管理）。<sup>1</sup>

#### （5）美国匹兹堡大学 David Bearman 的元数据系统

美国匹兹堡大学的 David Bearman 等人于 1995 年在其“保证电子文件凭证性的功能要求”的研究项目中提出的六层结构是目前国际上最为著名的文件元数据体系模型，并为澳大利亚国家档案馆等借鉴作为制订本国元数据标准的参照体系。David Bearman 的元数据模型

<sup>1</sup>刘琼瑶，国际档案界关于电子文件元数据之研究，四川档案，2003.2

包括六层结构：指示层、期限与条件层、结构层、背景层、内容层、利用历史层。<sup>1</sup>

指示层：文件说明、操作、领域、日期等标识数据。

期限与条件层：文档存取与使用权、身份、许可证、款项、密钥等，文件发表、引用、复制、保护、销毁等有关政策、法律、规章、机构、期限、区域等标识数据。

结构层：文档标识、鉴定、编码、字符、缘素、抽样、数据压缩、数据源（个人、系统、指令）算法、数据类型（光、声、温度、位置）、数据获取（个人、模式、背景、校准等）、特殊保护（隐私、专利、机密、版权等）、软件版本、硬件构成、加密、密钥等。

背景层：原作者、接收者、复制、使用、许可、契约、机构交易、授权、法律、组织、相关领域。

内容层：标准与类型的描述、文件、档案主体，包括可执行形式的信息。

利用历史层：使用类型、查看、复制、编辑、归档、编索引、分类、发送、处理等，价值、时间、用户标识数据。

#### （6）澳大利亚国家档案馆 Recordkeeping Metadata Standard for Commonwealth Agencies

1999年5月，澳大利亚国家档案馆公布了“政府记录保存元数据标准”（Recordkeeping Metadata Standard for Commonwealth Agencies），2008年7月，发布第2版。该标准建议澳

<sup>1</sup>刘琼瑶，国际档案界关于电子文件元数据之研究，四川档案，2003.2

大利亚联邦各政府机关的文件和信息管理人员、电子信息技术人员及其有关软件都应采用这一标准，以便各政府机关的鉴别、描述和管理文件与信息的方法能够做到系统化和保持一致，确保政府机关更加有效的管理文件和信息，以及公众更加便利地获取政府信息。<sup>1</sup>新版中定义了 26 个元数据元素，包括实体类型、分类、标识、名称、数据范围、描述、相关实体、变更历史、管理权限、安全等级、安全警告、权限、权利、联系信息、位置（由代理人持有的当前位置）、语言、覆盖范围、关键词、处置、格式、程度、介质、完整性检查、位置（该记录的当前位置）、文档形式、优先。26 个元素下还包括 44 个子元素。

2011 年 7 月，澳大利亚国家档案馆还发布了相应的指南（第 2 版），<sup>2</sup>指南给出了遵循“政府记录保存元数据标准”的最低要求，并为不同类型业务系统的记录保存元数据要求提供了建议。<sup>3</sup>

（7）美国印第安纳大学电子文件项目研究小组《文件管理元数据说明草案》

2000 年 7 月，美国印第安纳大学电子文件项目研究小组提出的《文件管理元数据说明草案》中，将元数据分为七个部分：标识、背景、内容、期限与条件、处置、结构、保护历史。<sup>4</sup>

---

<sup>1</sup>Australian Government Recordkeeping Metadata Standard Version 2.0【S】

<sup>2</sup>The Australian Government Recordkeeping Metadata Standard Implementation Guidelines.

<sup>3</sup><http://www.naa.gov.au/records-management/agency/create-capture-describe/describe/recordkeeping-metadata.aspx>.

<sup>4</sup>刘琼瑶. 国际档案界关于电子文件元数据之研究【M】四川档案, 2003(2): 34—35



标识：标识、类别、进入文件保管系统的日期、当前位置。

背景：发文机构、收文机构、文件发送日期、文件接收日期、相关文件之间的关系、机构职能及子职能之间的关系、可靠性。

内容：文件标题、文件创建日期、文件主题。

期限与条件：存取和使用权限、存取和使用的期限与条件、存取和使用历史。

处置：处置依据、处置时间、处置行为历史。

结构：数据格式、媒体格式、软硬件环境、保管标准、数据类型、文件集合水平。

保护历史：保护责任、保护活动发生的时间、保护措施的类型、储存介质。

#### (8) ISO 19106: 2004 Geographic information—Profiles

ISO 19106 Geographic information—Profiles (地理信息专用标准) 为制定 ISO 地理信息标准的专用标准提供了指南。ISO 19100 试图定义由 ISO/TC211 制定的 ISO 地理信息系列标准的专用标准概念，并且为制定这样的专用标准提供指南。只有满足该标准包含的专用标准定义的那些规范组成，才能通过该标准描述的机制进行建立和管理。这些专用标准能够使用 ISO 标准化程序达到国际标准化。ISO TC211 制定的 ISO 19100 系列标准为描述、管理和处理地理信息给出了各种模型、元素定义和规则，有些标准是抽象的不能直接实现，有些标准是相对通用的。同时，不同用户团体在使用或实现这些元素和规则的程度上有不同的需求。因此，ISO 19106 提出，为了明确确定和建立与

ISO 地理信息系列标准相一致的具体标准子集，需要专用标准，即需要与某一具体领域的特殊应用需求做进一步结合，派生出专用性较强、指导意义更为具体的标准。ISO19106 即提供了这方面的指南。<sup>1</sup>

#### (9) GB/T 18894—2002 电子文件归档与管理规范

GB/T 18894—2002《电子文件归档与管理规范》由国家档案局提出、起草并归口。《电子文件归档与管理规范》规定了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。标准适用于党政机关产生的电子文件的归档与管理，其他社会组织的电子文件管理可参照本标准。

标准中指出：“元数据指描述电子文件数据属性的数据，包括文件的格式、编排结构、硬件和软件环境、文件处理软件、字处理和图形工具软件、字符集等数据。”<sup>2</sup>

#### (10) GB/T 25100—2010 信息与文献都柏林核心元数据元素集

GB/T 25100—2010《信息与文献都柏林核心元数据元素集》由全国信息与文献标准化技术委员会(SAC/TC 4)提出并归口，由国家图书馆、上海图书馆、北京大学图书馆、清华大学图书馆、中国科学院国家科学图书馆、中国科学技术信息研究所共同起草。

本标准修改采用国际标准 ISO15836: 2009《信息与文献都柏林核心元数据元素集》，并参考了 ISO15836: 2003《信息与文献都柏林核心元数据元素集》、都柏林核心元数据计划(The Dublin Core

<sup>1</sup>ISO 19106: 2004, Geographic information—Profiles【S】.

<sup>2</sup>GB/T 18894-2002, 电子文件归档与管理规范【S】.

国家档案局  
www.saac.gov.cn

Metadata Initiative, DCMI) 发布的《都柏林核心元数据元素集》(1.1 版) (2008-1-14)、ANSI/NISO Z39.85: 2007《都柏林核心元数据元素集》(ISSN: 1041-5635) 及 RFC5013《都柏林核心元数据元素集》。

都柏林核心元数据元素集是应用于不同领域间资源描述的标准, 是根据各类资源的共同特点确定的元数据元素集合。本标准并不限定所描述的资源类型。<sup>1</sup>

(11) GB/T 26163-2010 信息与文献文件管理过程文件元数据第 1 部分: 原则

GB/T 26163.1-201《信息与文献文件管理过程文件元数据第 1 部分: 原则》由国家档案局提出, 全国信息与文献标准化技术委员会 (SAC/TC4) 归口, 由国家档案局、天津市档案局、安徽省档案局、江西省档案局、深圳市档案局、南京政治学院上海分院共同起草。本部分标准为 GB/T26163 的第 1 部分。本部分等同采用 ISO23081-1: 2006《信息与文献文件管理过程文件元数据第 1 部分: 原则》。

标准规定了文件管理元数据创建、管理及应用的原则, 适用于: 文件及其元数据、影响文件及其元数据的全部过程、文件及其元数据所处的各类系统、负责管理文件及其元数据的各类组织机构。标准提出的元数据类型包括: 关于文件自身的元数据、关于业务规章制度、方针和法规的元数据、关于责任者的元数据、关于业务过程的元数据、关于文件管理过程的元数据。<sup>2</sup>

<sup>1</sup>GB/T 25100—2010, 信息与文献都柏林核心元数据元素集【S】。

<sup>2</sup>GB/T 26163. 1. 201, 信息与文献文件管理过程文件元数据第 1 部分: 原则【S】。

国家档案局  
WWW.SAAC.GOV.CN

## (12) SDS/T 2111—2004 元数据标准化基本原则和方法

科技部非常重视科学数据共享的元数据标准化工作。根据科学数据共享标准化工作的规划，科学数据共享工程开展了元数据相关标准（包括元数据标准化基本原则和方法、元数据内容、元数据检索和提取协议）的研究。

为保证科学数据共享工程的元数据标准统一在同一个元数据标准框架内，制定了 SDS/T 2111-2004《元数据标准化基本原则和方法》。该标准由中华人民共和国科学技术部基础研究司提出并归口，由国家信息中心、中国林业科学研究院起草，以便指导各个领域制定各自的元数据标准（领域元数据标准、专用元数据标准）。该标准描述了科学数据共享工程中元数据标准的三个层次，以及各层元数据标准之间的关系、制定顺序、制定的基本原则和方法等内容。<sup>1</sup>《元数据标准化基本原则和方法》，开创了元数据标准化上位类标准制定的先河。

## 2、当前研究中对电子文件元数据类型的划分

元数据应用于不同领域且具有不同的目的，根据文献分析大致可以分为以下几类：

(1) 描述著录。用于对提供电子文件数据的单位进行详细、全面的记录。数据元素囊括内容、载体、位置与获取方式、制作与利用方法，甚至相关数据单元方面等，数据元素数量往往较多。

(2) 资源组织体系。其中包括对电子数据资源集合的基本描述

<sup>1</sup>SDS/T 2111—2004，元数据标准化基本原则和方法【S】。

和对资源集合的知识组织体系的描述。

(3) 确认和检索。进行电子数据信息检索和确认所需要的元数据资源，其数据元素相对简单。

(4) 资源保护与长期保存。元数据的数量随着电子文件的产生、使用和保管的过程而不断增加，因此对电子资源进行长期保存，首要的就是对元数据的保存，除与对资源进行描述和确认相关的元数据以外，还需要保存包括详细的格式信息、制作信息、保护条件、转换方式、保存责任等内容。

### 3、电子文件元数据的作用研究

通过文献分析，元数据的作用主要体现在以下方面：

(1) 集成管理。何嘉荪教授认为，要实现电子文件的集成管理与集成服务，必须以元数据为最基本和最重要的工具，并在此基础上提出了“基于元数据的电子文件集成管理与服务模型”。

(2) 集中与交换管理。黄蓓光、王申康提出了元数据管理的另一个策略，即通过建立元数据交换途径，使处于异构系统中的元数据可以用这种途径来完成交换管理，从而实现元数据的集成管理。这种元数据交换途径包括元数据桥和元数据交换标准。

(3) 分布对象管理。杨晓春提出了分布式环境下的信息合作模型，设计了以区域为单位的分布式对象管理体系结构，以及相应的元数据管理。

## （二）我国部分省市档案馆对于电子文件和电子档案元数据的研究、应用情况

课题组对江西省、辽宁省、天津市、青岛市等省市档案馆关于电子文件和电子档案元数据的相关规范进行了分析研究。江西省档案馆制定了文书类、录音类、录像类、照片类共 4 类电子档案元数据方案，经过历时 4 年多的试行并修改，元数据方案使用于江西省档案馆，市、县两级档案局馆和省直单位参照执行。辽宁省档案馆制订了文书类电子文件元数据标准（试行），开展了“省直机关电子文件接收与归档平台”项目。天津市档案馆制订了文书类电子文件元数据方案，并要求将元数据与电子档案一起移交。青岛市档案馆制订了文书档案、专业档案、照片档案、网页档案、录音档案、录像档案的元数据方案，并将元数据归档与电子文件归档工作一并开展。

这几个省市档案局馆关于电子文件和电子档案元数据的研究与应用的特点有：

### 1、制定标准接近统一

通过对江西、天津等 4 省市文书类电子文件和电子档案元数据方案分析可知，其制定标准过程中主要参考了 ISO 23081 和 DA/T46-2009 等相关标准，结合各省市电子文件管理实际加以完善。以文书类电子文件为例，主要思想是将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、实体关系元数据四个域，规定了每个域包含的元数据元素及其结构、来源、内容描述、形式特征、

权限管理、存储位置、电子属性、数字化属性、电子签名等容器型元数据，并对元数据进行分类表述。

## 2、设计电子文件全过程元数据管理

江西、天津等省市档案馆设计方案（标准）均规定了以文件为单位的文书类电子文件形成、办理、归档、移交、保存、利用的元数据项内容，对电子文件全过程管理的元数据设计、捕获和著录提出了要求。虽然对电子文件全过程各阶段元数据进行了设计，但是在对具体元数据描述中并未区分其在全过程中所处的阶段，也未对不同元数据进行分类管理。

## 3、元数据元素数量较多

各省市方案中对元数据元素的设定数量均比较多，以文书类档案元数据方案为例，江西省 132 项、辽宁省 84 项、天津市 63 项、青岛市 84 项。这种大量设置元数据元素的方式有利有弊。一方面可以更加全面的反应电子文件背景、内容、结构及其管理过程，但另一方面大量的元数据元素需要系统采集和人工添加，有可能会给实际应用带来不少困难。

## 4、元数据与著录项数据关系交叉

各省市档案馆在制定的元数据方案中一般都遵循了《文书类电子文件元数据方案》，除了包含电子文件形成过程中及电子档案管理过程中由计算机系统自动产生的、且不可由人工干预的元数据，还包括

了电子文件归档、电子档案整理、编目、保管、移交、利用等过程中形成的元数据。而这些数据中著录项数据占据绝大多数，也就是说，在电子文件和电子档案元数据中，很大部分与著录信息数据交叉。

### 三、电子文件和电子档案元数据研究在实际应用中存在的问题与分析

#### （一）电子档案接收在北京市的试点情况及其分析

##### 1、试点情况

北京市档案局于 2015 年初在北京粮食集团有限责任公司（以下简称京粮集团）和北京市审计局（以下简称市审计局）两个单位对文书类电子档案接收开展试点工作。两个试点单位移交的电子档案代表了两种类型。

京粮集团移交了 115 件档案，每件档案由一个记录了 16 项元数据的 Excel 文件和若干个 .XML 格式或 .HTML 格式、.GIF 格式、.PDF 格式的原文组成。这一类型档案，原文格式多样化，元数据不符合 DA/T 46-2009 中定义的 88 项元数据规范，元数据项等同目录字段项。

市审计局移交了 74 件档案，每件档案由一个记录了 16 项元数据的 Excel 文件和一个 .XML 格式的原文组成。这一类型档案原文格式单一，元数据符合国家档案局发布的 DA/T 46-2009《文书类电子文件元数据方案》，元数据项等同目录字段项。

表 1 试点单位移交电子档案情况

单位	内容	年度	数量	容量	方式	介质	格式
市审计局	电子公文	2012	74 件	8.37M	离线	移动硬盘	xml
京粮集团	电子公文	2012-2013	115 件	121M	离线	光盘	xml, pdf

在两家单位提交的 16 项元数据项中，包括：档案馆名称、全宗名称、立档单位名称、全宗号、目录号、年度、机构或问题、馆编文号、题名、主题词、文件编号、责任者、日期、文种、主送、件数等项目。

## 2、试点中发现的问题

### (1) 技术层面的问题

两个试点单位只提交了每份电子文件的各 16 项元数据，由于两个单位系统开发等原因，无法提取《文书类电子文件元数据方案》(DA/T46-2009) 规定的所有 88 项元数据。

电子签章技术是各电子签章技术开发公司专有技术，并非通用技术，因此无法移交和接收电子签章技术。

电子档案移交前由 doc 格式转换成 xml 格式，造成少数电子档案的排版样式发生变化，与原始生成时不一致，不能保证电子文件的原貌。但内容无变化。

单份电子档案由不同数据格式组成，如：正文为 DOCX 格式、附件为 PDF 格式。

国家档案局  
www.na.gov.cn

xml 格式电子档案仍然能够对原文信息进行添加、删除、复制等操作，影响真实性、原始性。

接收环境不完备。一方面，目前没有电子档案接收平台或系统，只能单一离线接收；另一方面，离线接收也需要电子档案管理系统、数据存储容量空间、基础软件、阅读（播放）软件等，并按照标准建立数据库，提供数据检验的工具软件。

## （2）管理层面的问题

电子文件元数据与电子档案移交目录内容大量是一致的，如档号、题名、文号、密级、页数、格式、保管期限等都要重复著录，增加工作量。

电子档案移交目录中“公民隐私”和“控制使用标记”不易掌握原则，无法科学填写。

电子档案“件”的概念不明确。是单一文件还是组合文件要明确。

电子档案的存储结构以及档号聚合层级有争议。是 4 层聚合（全宗-目录-案卷-文件）还是 3 层聚合（全宗-目录-文件）要明确。

移交介质应更加灵活，光盘、移动硬盘都应允许使用。由移交单位提供移交介质手续繁琐，质量不易控制，可考虑由接收方统一提供接收介质。

## （二）元数据理论研究在实践过程中出现的问题

从文献调研中可以看出国内元数据的研究已经从过去专注于元数据概念、功能、作用以及各种元数据规范的介绍和比较等层面上，

开始深入到元数据的管理技术和应用实践研究阶段，在这个过程中新的问题不断产生。结合档案界电子文件和电子档案元数据的研究现状，针对现有电子文件和电子档案元数据在实践应用过程中产生的问题进行分析，主要存在以下几个问题：

### **1、数字档案馆建设过程中，电子档案元数据的分类标准认识不统一**

一些档案学者认为并非所有与电子文件和电子档案相关的描述性数据均为元数据，只有电子文件形成过程中及电子档案管理过程中，由计算机系统自动产生的，且不可由人工干预的数据为元数据。包括公文批办单上的内容、档案整理编目形成的信息等由文件和档案管理工作形成的数据均不应称为元数据。但这种观点不符合信息界的通行观点。

而另外一些学者则认为电子文件形成和电子档案管理过程中产生的数据均为元数据，并且认为随着元数据的不断产生需要对元数据不断封装。但这样做一方面造成不必要的空间浪费，同时由于没有对元数据按照重要性和使用者进行分类，造成实际工作中很难落实。

### **2、电子档案的元数据集合的无序增长**

元数据集是规定和描述标准元数据集元素以及元素间相互关系的框架。但是目前在档案学领域，多数研究都认为，在电子档案元数据集里，各元素地位平等和同样重要，没有进行功能的区分，而将

各类与电子文件相关的信息全部归为元数据，使得元数据成为一个筐，什么数据都往里装。造成的恶果是元数据的泛滥，且不仅需要系统采集元数据，还要人工添加大量元数据，造成实际操作困难，且无法保障电子档案元数据集合的有效性。

### **3、电子档案元数据格式呈现多元化、复杂化的趋势**

随着元数据格式多元化、复杂化，如何实现电子档案元数据的标准化和不同格式元数据之间的整合以及互操作已经成为一个严峻的问题，使得电子文件元数据的捕获、收集、管理、存储难度大，而且随着元数据研究范围的不断扩大，深度不断加强，这个问题将变得更为严峻，难度也将更大。

综合上述问题分析，明确电子档案相关的元数据与管理流程相关的管理元数据的关系及定位，如何实现电子文件和电子档案元数据的科学分类，并筛选有效元数据集，实现档案元数据的有效捕获、收集、存储和管理是电子档案管理面临的重要课题。

### **（三）破解存在问题的理论依据**

#### **1、元数据是保证电子文件具有凭证作用的数据**

国外档案学者，InterPARES 项目的主持者露西安娜·杜兰蒂的文章中对“真实性”和“可靠性”做了如下阐述：电子文件的真实性是指，文件由特定机构使用安全可靠的系统软件生成，并没有被篡改或破坏。电子文件的可靠性则指文件的内容值得信赖。这两个概念的区

别在于，真实性是指文件的来源可信，而且文件自生成之日起没被改动过，可靠性是指文件生成之时，其内容与客观事实相符。<sup>1</sup>

在文件生命周期中，文件上的客观历史记录及保留的真切的历史标记是使文件、档案具备凭证作用的根本所在。相对于纸质文件的凭证作用是固定在一定的载体之上因而能够显而易见，电子文件的凭证作用则是通过电子文件元数据来实现的。电子文件的原始性，是指电子文件最初形成时所具有的，且在传输、存储等操作过程中未被改动过的信息形态，在认定电子文件原始性时，以内容的原始性为认定的唯一方法。

建立一个新的电子档案“原件”概念对于电子文件的原始性和证据性十分必要。基本要点是要确认一份文件确实是当时由作者撰写或制作出来的，而不管其载体形态和现在格式如何。电子文件的凭证属性是由原始性承载的。电子文件中的信息的易变性使人们对其原始性产生不确定的判断，特定的信息不再固定于特定的载体，无法通过载体来判断它的原始性。<sup>2</sup>

电子文件具有客观真实性。凭证的客观真实性是指它所反映的内容是客观存在的事实，是法律关系发生、发展或变更、终止过程中形成的事实，或者是客观事实的真实记载与反映。电子文件具有特定的文体、特殊的语言体系、规范的结构和格式及特定的处理程序，这是保证文件内容真实性、有效性、正确性、完整性的重要条件。电子文

---

<sup>1</sup><http://www.interpares.org>

<sup>2</sup>原艳丽，电子文件凭证属性浅析，兰台世界，2008,3，下半月

国家档案局  
www.saac.gov.cn

件的撰写、传递、处理、审核无不遵循这些规范、格式，从形式和程序上保证了自身内容的真实性。从内容上说，只要一份电子文件的内容确实是在当时由原作者撰写或制作出来的，此后又从未被修改过，就具有客观真实性，尽管它无固定载体，没有实在的物理形态，甚至因转换而失去了原来的格式。这种真实性概念与纸质文件的区别在于，它抛弃了文件形式上的原始性，仅以文件中所含信息的真实、准确，即内容实物原始性为唯一标准，这是电子文件具有凭证作用的基础。

1

通过电子文件元数据即可以判断电子文件归档后形成的电子档案的内容是否与形成时的电子文件完全一致，从而证明电子档案的真实性和完整性。电子文件元数据中包含着电子文件形成的时环境和背景信息，是证实电子文件初始状态的依据，也是归档后电子档案凭证价值的依据。电子档案具有凭证性，是由电子文件的原始性承载的，电子文件元数据保障了电子文件中的各类信息的不变。

## 2、文件连续体理论

澳大利亚档案学者 Frank Upward 于 20 世纪 90 年代提出了一种全新的“文件连续体理论”，从多维、立体的角度对电子文件生命周期作了描述，即“往复运动从生成到处置的连续体中的一个过程”<sup>2</sup>。该理论构建了一个包括四个坐标轴的多维体系，将电子文件的生命运

<sup>1</sup>吴坚，试论电子文件的凭证作用，兰台世界，2010.10 下半月

<sup>2</sup>Frank Upward, Structuring the records continuum. Archives and Manuscripts. Vol,24. No.26

动过程及其与外界的联系清晰的展现出来。如图 1 所示。可以看出文件保管轴是核心轴，它的变化带动了其他坐标轴的变化，而第一围则突出表现了文件生成的过程。

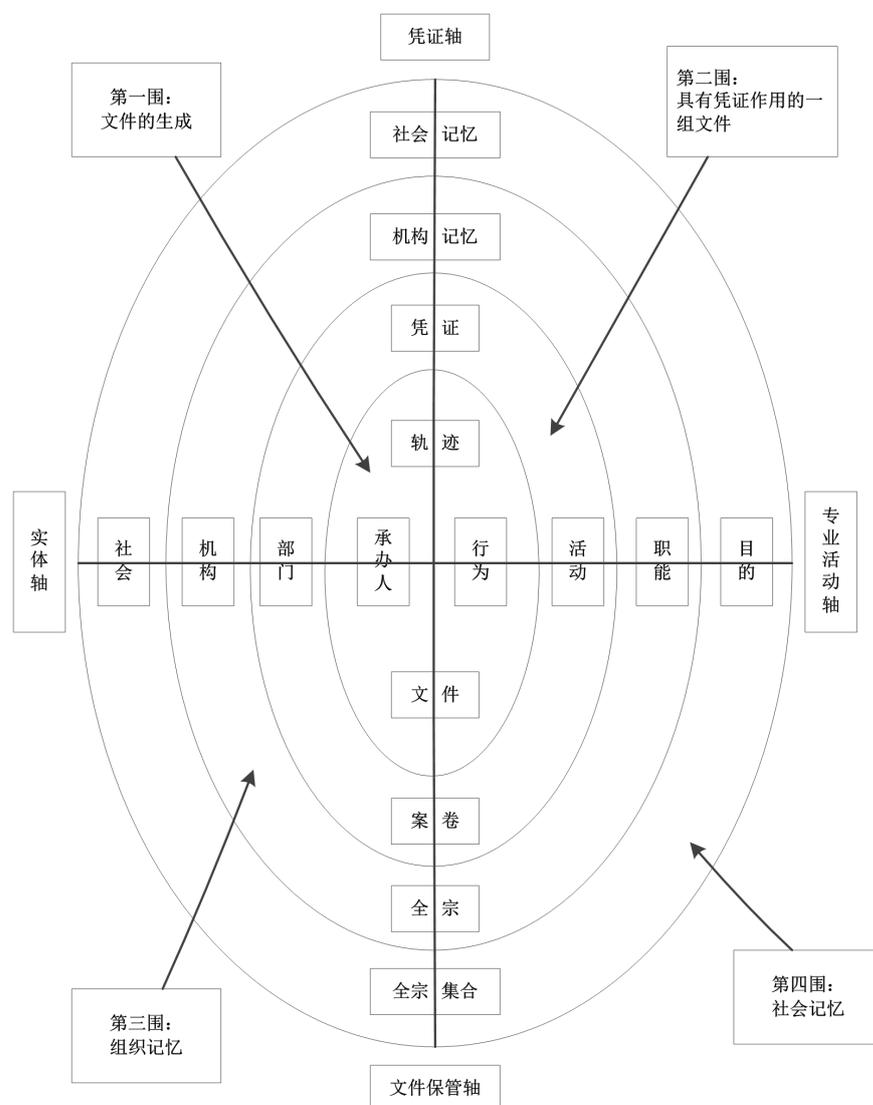


图 1 文件连续体模型

### 3、“前端控制”思想

图示指出“档案工作者干预电子文件的时机，应确定在决定必须保存文件和设计归档系统阶段”<sup>1</sup>，把需要和可能在文件形成阶段实

<sup>1</sup>International Council on Archives, “Guideline on Electronic Records Management”, 1997年2月

现或部分实现的管理功能尽量在这一阶段实现，使一些传统档案工作中的后工序出现的问题提前介入，以确保有价值的电子文件不丢失。

#### 四、电子文件和电子档案元数据分开管理的设想

##### （一）分开管理的目的和原则

###### 1、明确使用电子文件和电子档案元数据的目的

保证电子文件的真实性是使用电子文件元数据的根本目的，进而通过电子文件元数据确保档案数字资源的真实性和凭证作用。区别于其他学科中涉及到的各类元数据，电子文件元数据有着自己独特的属性和内容。电子文件元数据的使用，和相近学科元数据的使用有很大不同，例如与图书、文博、情报等学科的元数据，在使用上有着本质区别。

区分电子文件各类相关数据的功能，在相关学科中使用元数据，用以反映相关资源实体描述对象、形式、业务、责任者关系、保存等内容，突出了信息描述、信息整合、信息检索和信息共享等作用。

电子文件在生命周期中，以至于归档后形成档案数字资源的存在过程中，会形成多种类型的数据，电子文件的各类相关数据主要分为两类，一是说明电子文件本身的数据、各种原始痕迹的数据，即电子文件元数据，反映的是电子文件的原始与真实。二是电子文件以及电子档案的管理、使用迁移过程中的数据，反映的是电子文件及电子档案的各种管理。

## 2、明确保证电子文件和电子档案真实性元数据的本质特征

### (1) 依赖系统生成，伴随电子文件的产生而产生

电子文件具有系统的依赖性，从电子文件创建、形成、定稿、运转、处置、归档、形成电子档案的全过程都必须依赖于系统来实现，所形成的各种痕迹、特征是通过电子文件元数据忠实记录下来。元数据的形成是对系统依赖的直接体现。元数据是使电子文件信息完整的重要组成。这些痕迹、特征的数据是系统创建之初，根据功能需求而由计算机系统自动实现的，而非人为增删改而出现的。

### (2) 不可更改

电子文件制作的本身就是数字化、虚拟化的过程，在客观上造成“无法通过固定载体来判断是否是原件”的现象。由于电子文件的数据性，电子文件的形成、运转等都是实时的，通过严格的收集、提取、存储等程序，避免使其失真因素的出现，一经形成便始终保持最初、最原始的状态，能够客观反映文件真实的本来面貌。而管理信息则是具有可变性的，管理信息是动态中的，数据处于不断更新和补充中，同时具有可操作性。

在我国《电子文件元数据标准》中，给出了电子文件元数据的基本描述，就是描述电子文件内容、背景和结构信息及整个管理流程的数据，本质是描述对象的各种属性及相关关系等内容，每一个元素的语义明确定义了其含义、内涵与外延。语义定义是一个元数据规范的

基本要素，也是元数据应用与元数据互操作的最重要的基础。电子文件形成、捕获、登记、分类、存储和保管、利用、跟踪、处置、传输、归档移交及长期保存等过程中都需记录在元数据中，并应保持连续、一致，以确保电子文件的真实性、完整性与有效性，实现电子文件全过程管理与监控，电子文件及其元数据与结构稳定。

在此前提下，提出了很多的元数据分类方法，一般根据功能划分元数据的类型，如表 2 所示。

表 2 元数据类型

类型	定义
描述型元数据	用来描述或识别信息资源的元数据
管理型元数据	在管理信息资源中利用的元数据
保存型元数据	与信息资源的保存管理相关的元数据
技术型元数据	与信息系统如何运行或元数据如何发挥作用相关的元数据
使用型元数据	与信息资源的注册和类型相关的元数据

## （二）电子文件和电子档案元数据划分为对象元数据和事件元数据

将电子文件元数据分类就是针对电子文件和电子档案的性质、特点、用途等作为区分的标准，将符合同一标准的元数据聚类，不同的则进行区分。

档案界的元数据与其他行业元数据有很大不同。电子文件和电子档案与图书等类最本质的区别在于电子文件和电子档案的原始凭证

性和真实性，而在考虑电子文件和电子档案元数据时最重要的标准就是保证电子文件和电子档案的原始凭证性和真实性的元数据与其他元数据的区别。

与电子文件伴生的、可以说明电子文件性质的元数据，是确保电子文件真实性的重要、核心元数据，包括电子文件产生时的电子文件自身属性、技术环境等。

其他类型元数据包括对电子文件归档、在档案室和档案馆进行档案管理产生的具有管理属性的元数据，涉及电子文件归档、电子档案移交和接收工作、电子档案的鉴定和解密划控工作、电子档案的迁移工作的管理信息，以及电子档案的利用信息等。

根据电子文件的类型将其划分为保证电子文件真实性、原始性的元数据，以及在电子文件管理、归档、长期保存和利用过程中形成的元数据。

从电子档案元数据的使用途径角度进行分类，可以分为保障电子档案真实性的“对象元数据”和支持电子档案管理的“事件元数据”。

结合档案领域的实际工作情况，认为表 2 中提到的描述型元数据是电子档案元数据的核心，可以称为对象元数据，这类元数据具有伴生性，由系统自动生成，伴随电子文件的产生而产生且不可更改，具有原始、凭证、真实性的特征，可以用于描述或标识电子档案资源的内容和外部特征，通过这些描述型元数据，用户可以得到电子档案文件的主题、来源等方面的信息，以全面了解电子档案资源，进而对电子档案进行评估。

而管理型元数据、保存型元数据、技术型元数据和使用型元数据中大部分元数据信息都可以被认为是对电子档案管理起辅助作用的元数据类，可以称为事件元数据，是在电子档案文件的管理、迁移和利用过程中形成的，服务于电子档案文件日常维护和管理。

对象元数据的核心是印证电子档案真实性，而电子档案真实性有两个阶段组成，一是形成过程中的真实性，二是归档后长期保存过程中的真实性。其中，第一个阶段的真实性是电子档案较为核心的部分，其关系到电子档案的法律地位。而第二个阶段的真实性重点需要关注的是防篡改，也就是第二个阶段重点需要确保第一个阶段的真实性没有受到破坏，这也是档案部门的重要职责。

电子档案身份证技术就是在第二阶段中确保电子档案的核心属性（真实性）未被破坏，可通过严谨的技术手段保证电子档案在长期保存过程中的法律地位。

课题目前研究的设计是：电子档案本身不再与元数据做封装，但电子档案身份证中包含对象元数据、内容信息的全文哈希值和身份证号等内容，并做加密处理。电子档案身份证中只绑定电子文件形成过程中、与电子文件真实性直接相关的对象元数据。后期在档案室、档案馆产生的部分其他元数据就不与电子档案身份证绑定在一起，而是在数据库中以结构化数据形式保管。

因此，对象元数据是那些在电子文件形成过程中产生的，记录了电子文件形成过程中重要事件的项目，且其在定稿后，按照制度是不允许再做任何修改的那部分元数据。

### （三）概念界定

#### 1、对象元数据

将一组数据与这组数据有关操作组装在一起，形成一个实体，这个实体就是对象。在应用领域中有意义的、与所要解决的问题有关系的任何事物都可以作为对象，它既可以是具体的物理实体的抽象，也可以是人为的概念，或者是任何有明确边界和意义的东西。

电子档案对象元数据是指电子文件形成过程中产生的，由信息系统自动生成且不可人工干预的，直接关系电子文件和电子档案凭证性的元数据项，对象元数据应封装在电子档案身份证中。

这类元数据是电子档案元数据的核心，它具有伴生性，由系统自动生成并且不可更改，也包括事先由人工在电子文件形成系统中进行预置，在电子文件产生时由系统根据流程设计和智能判断自动附加在电子文件上的元数据。这类元数据可以用于描述或标识电子档案的内容和外部特征，可以反映电子档案的主题、来源等方面的信息，可以证明电子文件形成过程真实、办理过程真实、归档过程真实，确保电子档案凭证价值的发挥。

#### 2、事件元数据

在计算机技术中 visual basic 的窗体和控件是具有自己的属性、方法和事件的对象。可以把属性看作一个对象的性质，把方法看作对象的动作，把事件看作对象的响应。事件是可以被控件识别的操作，

如按下确定按钮，选择某个单选按钮或者复选框。每一种控件有自己可以识别的事件，如窗体的加载、单击、双击等事件，编辑框（文本框）的文本改变事件，等等。

电子档案事件元数据是指电子文件归档和电子档案管理过程中产生的，对电子档案溯源起重要作用的，由信息系统自动生成或人工录入的元数据项，事件元数据以结构化的数据形式存储在数据库中，并通过电子档案目录与电子档案的内容信息和电子档案身份证进行有效关联。

事件元数据是电子文件和电子档案的技术类属性，以及在电子档案长期保存过程中对其进行管理产生的，附加给电子档案自身的属性和管理事件、行为的记录。管理型元数据、保存型元数据、技术型元数据和使用型元数据中大部分元数据信息都可以被认为是对电子档案管理起辅助作用的元数据类，可以作为事件元数据，在电子档案文件的管理、迁移和利用过程中形成的，服务于电子档案文件日常维护和管理。

## 五、电子文件和电子档案元数据分开管理系统框架

考虑到档案领域数据管理的行业性、特殊性，本课题通过研究认为电子档案元数据的管理主要从保证电子文件和电子档案凭证性、真实性和满足正常电子文件和电子档案管理流程两个方面进行规划和设计。在此基础上，提出电子档案元数据分开管理的概念，结合电子文件和电子档案的形成与管理环境特点，从两个角度对可能产生的元

数据分析，即与电子档案真实性直接相关的元数据类和对电子档案管理起辅助作用的元数据类，设计电子档案元数据分开管理系统框架，其设计流程如图 2 所示。

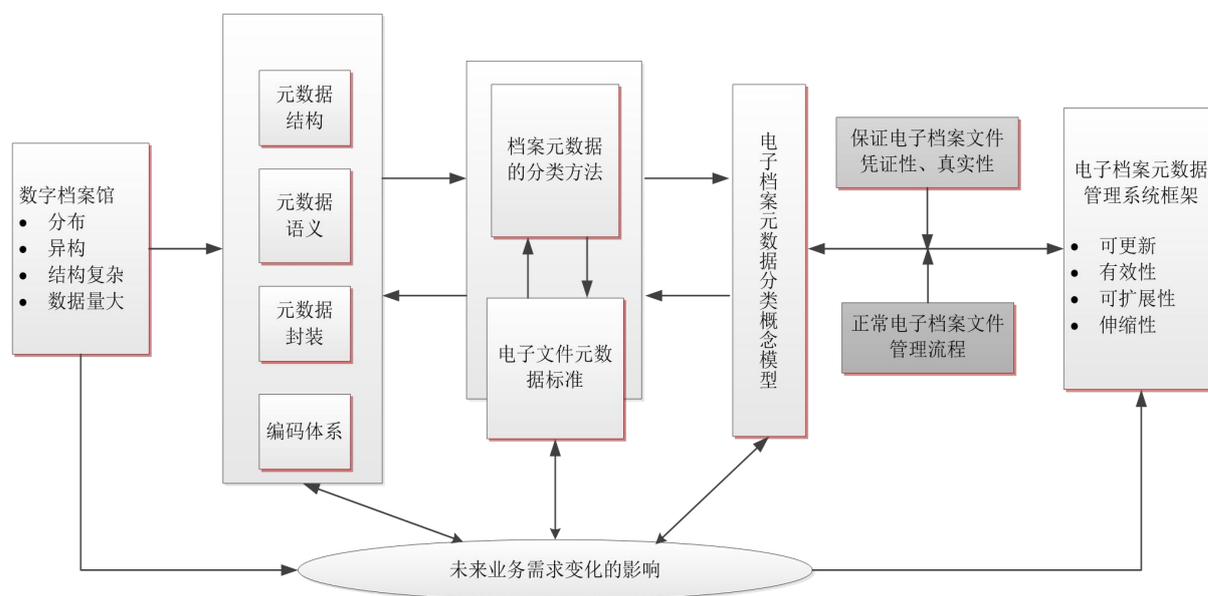


图 2 元数据管理框架的设计流程

分开管理的最终的目标是在电子档案元数据管理中能够实现在“正确”的时间，在“正确”的场景，选择“正确”的元数据，给“正确”的人，从而提高电子档案元数据管理的效能。当能有效的管理元数据时，数据才能变得更有价值。

### （一）电子文件和电子档案元数据分开管理系统框架的特点

元数据作用的发挥是在元数据管理系统中实现的，从更大的范围来看是在电子文件和电子档案管理系统中实现，所以元数据系统是电子文件和电子档案管理系统中的子系统，是电子档案管理系统中重要

的有机组成部分，应具备如下特点：

### **1、立足档案管理需求，划分档案元数据分类并进行分别管理**

结合电子文件和电子档案的形成与管理环境特点，对全部可能产生的元数据进行科学分类，分类过程中重点考虑两类：即与电子档案真实性直接相关的元数据类，和对电子档案管理起辅助作用的元数据类。即从电子文件形成至电子档案长期保存全生命周期中产生的全部元数据项划分为两大类，即对电子档案真实性保障起决定作用的“对象元数据”，以及可以为电子档案管理工作提供辅助的“事件元数据”等。

其中，对象元数据包含电子文件形成过程中产生的业务背景和法律背景，事件元数据则主要指电子文件形成过程中产生的技术背景和电子文件归档以后产生的管理过程。通过课题研究，我们提出，对象元数据将与电子档案身份证捆绑在一起，事件元数据则不需要与电子档案身份证绑定。

### **2、提供更加有效的元数据集合**

参考国内外档案和图书行业在电子公文、数码照片和数字声像文件等电子文件类型的元数据标准，结合我国档案数字资源的分布特点，对以上几种电子文件类型进行元数据设项分析，筛选并建立符合电子档案管理所需的元数据总集，即从电子文件形成阶段开始梳理，直至电子档案长期保存的循环性管理阶段为止，对可能产生的全部元数据

项查全、找准，再与著录信息项对比、去除冗余数据项后，明确其在形成到长期保存过程中应记录和保存的全部元数据项，建立电子档案的元数据总集，并对各元数据项进行解释。随后，根据电子档案真实性、凭证性的保障要求，遴选电子文件在形成过程中最直接反映其真实性的核心元数据子集，即有效的、具有凭证性的元数据集合。

### 3、支持实现元数据集合的质量控制

由于元数据影响着电子文件和电子档案的原始性、真实性和凭证性，所以元数据的质量控制也是非常重要的。有关文献<sup>1</sup>曾将元数据质量定义为“元数据满足需求和目标的程度”，具体到电子档案元数据则是指元数据描述档案信息资源的完整性、准确性及其自身的长期保存性，减少元数据的遗漏、错误，减少元数据冗余，增强元数据之间的互操作性。

元数据的完整性要求对于电子文件的内容、背景、结构及过程进行详细全面的描述，减少元数据的遗漏，在设计元数据管理系统之时就需要将相关采集项、输入格式等编制完整。在保证完整性的基础上还要保证元数据描述的精度，描述的精度是指元数据对数字对象描述的准确度，其描述的行为与真实值之间的差异。减少元数据错误，这种错误一是由于在系统设计之初所造成，及对于元数据标准的理解错误而造成编码的错误致使系统不能正常的运行；二是人为输入造成的。元数据形成一般有两种方式，即：由系统自动分析提取所需数据和由

<sup>1</sup>刘家真，廖茹. 电子文件管理元数据的质量控制与管理[J]. 图书情报知识, 2009(132).

人工输入完成。在人工输入时有很大的机率会造成字符录入错误，为日后的管理与检索带来不便。再者就是要降低元数据的冗余，造成这种现象的原因也主要集中在系统设计之初，由于在选择元数据元素时缺乏与实际应用部门的沟通与协调，对于有些元数据元素不加选择的编入系统之中，从而造成元数据的过量采集与存储，造成存储空间的浪费，也会给元数据之间的互操作带来困难。

## （二）分开管理元数据的分类标准

具体到电子文件和电子档案元数据，哪些属于对象元数据、哪些属于事件元数据，以什么为分类管理的原则和标准，是本课题重点研究的内容。其原则包括：

### 1、实现元数据管理方便、操作便捷的目的

电子文件和电子档案元数据分开管理的目的是要对档案的管理方便、操作便捷、管理实用性强。

通过对相关元数据标准和规范的理论梳理和实践分析，将对于实际工作中可能涉及的不明确、不清晰的元数据内容重新清晰化，并将其较为繁琐的内容更加直观化，更易于操作，且使在系统建设初就明确定位。元数据分开管理后，要达到的目的是：符合档案业务工作流程需求，保证电子档案真实可靠，符合系统使用需求。

### 2、元数据划分类型逻辑性强并与著录信息区分

基于理论和实践研究，将元数据分为对象元数据、事件元数据，

保证电子档案真实性的元数据为对象元数据，支持电子档案管理的元数据为事件元数据。而这两类元数据分开管理后，其反映电子档案真实性的一些元数据作为电子档案身份证的一个项目组成部分与电子档案身份证绑定。而另外描述电子档案的各种著录信息将作为著录项在数据库中保存。

### 3、让不同使用对象认识更明确、操作更简单

元数据的概念对于大多数档案管理人员、特别是基础档案室的档案管理人员来说，是比较生涩的概念。分开管理的目的之一就是要使各工作节点、部门的相关档案管理人员在文件档案管理全过程中，能够对电子文件和电子档案元数据简单化操作。因此，分开管理的分类要明确在使用过程中的不同使用对象，由系统设计人员在系统研发过程嵌入系统，哪些是在自动化办公系统中进行预置、自动保存，哪些是档案室人员记录，哪些是档案馆人员记录及使用。将元数据保存与使用根据简单化，将更加易于保证电子文件和电子档案的原始性、真实性和凭证性。

#### （三）分开管理的具体方案

本课题对元数据分类管理研究中，分别重点研究了文本类、图像类、影像类、声音类 4 类电子文件和电子档案的对象元数据和事件元数据。

国家档案局  
www.saac.gov.cn