



中华人民共和国国家标准化指导性技术文件

GB/Z 23283—2009/ISO/TR 18492:2005

基于文件的电子信息的长期保存

Long-term preservation of electronic document-based information

(ISO/TR 18492:2005, IDT)

2009-03-13 发布

中华人民共和国国家质量监督检验检疫总局 发布
中国国家标准化管理委员会

前 言

本指导性技术文件等同采用 ISO/TR 18492—2005《基于文件的电子信息的长期保存》(英文版)。

本指导性技术文件与技术报告 ISO/TR 18492—2005 在内容上相同,但做了以下编辑性修改:

——删除 ISO/TR 18492—2005 的前言,编写本指导性技术文件的前言;

——删除 ISO/TR 18492—2005 的目录;

——在术语和定义增加 GB/T 20225—2006;

——改正了附录 A.3 中出版物序号的错误。

本指导性技术文件的附录 A 为资料性附录。

本指导性技术文件由全国文献影像技术标准化技术委员会(SAC/TC 86)提出并归口。

本指导性技术文件由全国文献影像技术标准化技术委员会七分会负责起草。

本指导性技术文件主要起草人:张美芳、孙静荣、李娟、姜志伟。

引 言

确保真实基于文件的电子信息的长期保存在档案学、文件管理、电子商务、电子政务和技术开发等许多领域都是突出的问题。为解决这一问题,长期保存基于文件的电子信息的个人和机构已采取多种方法和措施。

显然我们需要解决真实的基于文件的电子信息的长期存储问题,但在解决这些问题上,目前还缺乏一个统一的国际标准,这就导致所采用的方法不一致甚至不兼容,有可能直接影响基于文件的电子信息的可用性和真实性。基于计算机硬件和软件的普遍存在的技术过时现象和数字存储介质的有限寿命等,这份指导性技术文件为保存机构利用和保存基于文件的电子信息提供指导性建议。

本指导性技术文件为基于文件的电子信息长期保存策略的制定和最佳实践提供了清晰的框架。使它能广泛地宜用于公共部门和个人,以确保基于文件的电子信息的长期可用性和真实性。

基于文件的电子信息的长期保存

1 范围

当基于文件的电子信息的保存期超过产生和维护信息所使用的技术(硬件和软件)的预期寿命时,为长期保存和检索真实的基于文件的电子信息,本指导性技术文件提供了实用的、方法性的指南。

本指导性技术文件考虑了中立的信息技术标准在信息长期利用中的作用。

本指导性技术文件同样认为,为确保基于文件的电子信息的长期保存和检索,需要信息技术专家、文件管理者和档案工作者的共同努力。

本指导性技术文件没有涵盖真实的基于文件的电子信息的产生、获取和分类的全过程。

本指导性技术文件适用于由信息系统产生的作为业务活动凭证的所有形式的信息。

注:基于文件的电子信息构成了日常业务活动或事件的“业务记忆”,使得机构日后可以回顾、分析或证明这些活动和事件。基于文件的电子信息是业务活动的凭证,可以帮助机构现在或将来做出管理决策,满足用户的需求和应对不利的诉讼,为此,基于文件的电子信息宜被适当地保留和保存。

2 规范性引用文件

下列文件中的条款通过本指导性技术文件的引用而成为本指导性技术文件的条款。凡是注日期的引用文件,其随后所有的修改单(不包括勘误的内容)或修订版均不适用于本指导性技术文件。凡是不注明日期的引用文件,其最新版本(包括修订本)适用于本指导性技术文件。

GB/T 20225—2006 电子成像 词汇(ISO 12651:1999, IDT)

ISO 15489-1 信息与文件 文件管理 第1部分:概论

ISO/TR 15489-2 信息与文件 文件管理 第2部分:指南

ISO/TS 23081-1 信息与文件 文件管理过程 文件的元数据 第1部分:原则

3 术语和定义

GB/T 20225—2006、ISO 15489-1 和 ISO/TR 15489-2 确立的以及下列术语与定义适用于本指导性技术文件。

3.1

真实的基于文件的电子信息 authentic electronic document-based information

准确性、可靠性和完整性得以长期保持的基于文件的电子信息。

3.2

基于文件的信息 document-based information

能作为一个整体来处理的独立存在的信息(例如一个影像、主题、电子表格、数据库视图等)。

注:基于文件的信息包括,但不必局限于文本、影像、表格数据(如电子表格)或者其组合。

3.3

基于文件的信息内容 document-based information content

包含于基于文件的信息中的实质内容。

3.4

基于文件的信息背景 document-based information context

有关电子文件的产生、控制、使用、存储和管理的信息,以及与其相似材料的有关信息。

3.5

基于文件的信息结构 document-based information structure

基于文件的信息的逻辑属性和物理属性。

注：逻辑属性包括逻辑顺序，如可区分子部分的层次结构，而物理属性包括各种要素，如字体、间距。

3.6

电子存档 electronic archiving

以一个独立的物理或逻辑空间存储电子信息，以免丢失、更改或破坏。

注：如果以这种方式保护信息，也许将来会被当作可靠的凭证。

3.7

长期保存 long-term preservation

基于文件的电子信息保持可用性和真实性的期限。

注：这个时间可能是几年至几百年，主要取决于机构的需要与要求。对于一些机构，保存时间由管理者的需要、法定需要和业务需要而决定。对于档案馆这样保存公共文件的机构，保存基于文件的电子信息的期限通常为几百年。

3.8

元数据 metadata

描述基于文件的电子信息内容(包括检索的索引术语)、背景和基于文件的电子信息结构以及长期管理的数据。

3.9

迁移 migration

几乎不改变基于文件的电子信息的结构、内容和背景，将信息从一种硬件或软件环境或存储介质转移到另一种环境或存储介质的过程。

3.10

保存机构 storage repository

承担真实基于文件的电子信息存储与保存任务的保存机构或实体。

注：这种定义不同于“保存机构”的技术定义。

3.11

技术过时 technological obsolescence

在业内，由于技术的发展和进步而采取的对已有技术方法的取代。

4 代号与缩略语

ASCII	美国信息交换标准代码
CRC	循环冗余码校验
HTML	超文本链接标识语言
JPEG	联合图像专家组(一种压缩标准)
OCR	光学字符识别
PDF/A-1	可移植的文档格式
SHA-1	标准散列算法
TIFF	标记图像格式
WORM	一次写入多次读取
XML	扩展标识语言

5 长期保存

5.1 概述

随着计算机信息的产生、利用、存储和保存技术的日益发展,私人和公共部门越来越依赖基于文件的电子信息作为其业务活动的正式凭证。保存机构面临着如何确保真实的基于文件的电子信息长期保存的挑战。信息产生于安全可靠的信息系统,存储于电子介质中,二者可能面临技术过时的风险,如果因技术过时而遗留没有纠正的错误,将会导致基于文件的电子信息无法恢复。各个机构的活动和业务逐渐在无纸化的环境中进行,这一现实加深了“确保真实基于文件的电子信息长期利用”的重要性。

因此,制定和应用长期保存及检索真实基于文件的电子信息的明确策略是非常必要的。5.2 条款涉及到策略的内容。

5.2 长期保存策略的目标

5.2.1 概述

本条款明确了保存机构制定保存策略时宜考虑的六项关键要素。

5.2.2 可读的基于文件的电子信息

长期保存策略宜确保基于文件的电子信息在未来可读。为了实现此目标,构成基于文件的电子信息的比特流在以下任何过程可在计算机系统或设备上利用:

- 文件生成时;
- 当前存储时;
- 当前利用时;
- 将来被用来存储电子信息。

这四个过程中都有可能使存储于介质中的基于文件的电子信息变得不可读。这主要由两种原因造成。

一是不适宜的保存环境。目前用来存储基于文件的电子信息的所有介质普遍容易受到不适环境的侵害,如温度、湿度的波动。这些不利条件会毁坏介质或加速其老化过程。不同类型的数字存储介质需要不同的保存环境以确保其最大寿命。一些存储介质在磁场干扰、灰尘和环境污染物的作用下容易产生信息损坏(磁性存储介质),而其他不易受到外界因素影响的介质(光学存储介质),只要稍加控制存储环境,介质就不易受损。不管使用何种存储介质,关键是要意识到不利的环境会引起各种形式的存储介质的损毁或老化。

二是介质过时。介质过时也会导致信息无法读取。存储介质(如磁带或光盘)与使用的计算机硬件(如磁带或光盘设备)存在物理上不兼容,导致信息无法读取。由于信息技术的不断发展,介质今后过时是不可避免的,因为储存技术的进步会继续导致基于文件的电子信息物理存储方式的改变(如记录技术、磁盘驱动硬件/软件接口的改变),存储介质形式和表现记录信息的比特流是一种物理形式(如纠错码)。因此,随着时间的推移,存储介质将会与以后使用的介质不兼容。

因介质过时,长期保存策略宜特别重视将基于文件的电子信息定期从旧的介质转移到新的介质上。

注:数据的可读性与数据的格式同样重要。宜考虑确保数据格式(技术中立格式)将来能被用户处理。

5.2.3 可理解的基于文件的电子信息

长期保存策略宜提供可理解的基于文件的电子信息。如果计算机能够描述如何解释比特流,数字信息才能被计算机理解。因此,基于文件的电子信息的可理解性是关于比特流实际代表信息的功能和根据这种信息采取适当措施的能力。

示例:构成数字标记图像格式的二进制代码本身不具有可理解性。然而,影像文件的页眉,通过使用字节顺序和压缩法能使计算机(通过操作系统和影像软件的结合)显示和打印影像。同样,文字处理文件携带元数据,使得文字处理软件易于理解其内容。

5.2.4 可识别的基于文件的电子信息

长期保存策略宜提供可识别的基于文件的电子信息。宜按特定方式组织、分类和描述可识别的基

于文件的电子信息,使用户和信息系统能够依据名字或 ID 号码等唯一属性区分信息对象。按照共享的属性把基于文件的电子信息归于各种类别以便查询和检索。无法提供这样的识别方式会严重限制查询和检索。

5.2.5 可检索的基于文件的电子信息

长期保存策略宜提供可检索的基于文件的电子信息,意味着能检索和显示离散的信息对象(或一部分)。可检索性通常依赖于软件,它需要特定的键值或指针将信息对象的逻辑结构(如数据域或字符串)与物理存储位置连接起来。

一般来说,在数据库记录、文件系统目录结构、文件分配表、文件头或标签中有这种链接。它包括以下信息:一个对象开始的位置信息、每个要素或数据要素的字节数目和它们在存储介质上物理位置等。

基于文件的电子信息的逻辑结构是一个操作系统功能或结合特殊的应用系统的设备驱动器,这个应用系统可用于存储、管理和利用数字信息。因此,信息对象的可检索性不可避免地及设备驱动器、应用软件、文件系统或操作系统联系起来。

与旧文件格式兼容的新一代文件格式有助于确保基于文件的电子信息的可检索性。然而,向后的兼容性可能会受到限制,因为许多软件开发商只支持特定的文件格式,而其他软件开发商支持所有版本的不同数据格式。例如超文本链接标示语言、静态图像压缩标准、标记图像格式都具有向后兼容性。

5.2.6 可理解的基于文件的电子信息

长期保存策略宜保证基于文件的电子信息可以理解。可理解的基于文件的电子信息宜把信息传递给计算机和人类。然而,分散的基于文件的电子信息不仅由内容决定的,而且由产生和使用的背景获得的(如元数据)。因此,保存机构宜意识到确保基于文件的电子信息的可理解性与确保纸质文件的可理解性有很大不同。纸质文件物理特性很典型地表达出它们产生和使用的背景,而产生和使用基于文件的电子信息的背景通常存在逻辑上的联系而不是物理上的。

示例:关于一个特定活动的纸质文件也许被装订在一起或放置在一个文件夹里。而一个类似活动的基于文件的电子信息也许存在于不同位置的不同介质中,因此,宜通过电子形式联系在一起。这些逻辑上的联系可以包括有关业务处理过程及参与者的证明。

产生和使用基于文件的电子信息的背景也包含以各种方式获得的基于文件的电子信息之间的关系,包括在文档描述中有关用其他材料处理同样问题的参考代码,或连接每一份与相似处理有关的基于文件的电子信息的分类代码。

无论保存的时间长短,基于文件的电子信息的成功检索部分取决于这些逻辑关系的保存。

5.2.7 真实的基于文件的电子信息

5.2.7.1 概述

长期保存策略的关键目标是确保真实的基于文件的电子信息。真实性是基于文件的电子信息保存的目的所在,如随着时间推移,可靠的信息没有被改变、更改或破坏。试图提供可长期利用的真实基于文件的电子信息的机构宜在他们的策略中考虑以下三个关键的措施:

- a) 迁移和保存;
- b) 存储环境;
- c) 利用和保护。

5.2.7.2 基于文件的电子信息的迁移和保存

只要电子文件保留在产生环境中,而没有储存在无法更改的一次写入式的介质上,保证其免受更改是很困难。长期保存策略宜提供基于文件的电子信息从其生成环境、产生者和接收者到保存系统或保存库房的迁移措施。比如根据文件的政策和实践,由独立运作的第三方负责保存基于文件的电子信息。

5.2.7.3 存储环境

长期保存策略宜明确说明存储基于文件的电子信息介质所需一个稳定的存储环境。因为,不利或不适宜的存储环境使基于文件的电子信息处于危险中。

5.2.7.4 基于文件的电子信息的利用和保护

长期的保存策略宜提供限制利用基于文件的电子信息和免受蓄意和意外更改和破坏的方法。

存储在可擦写式介质上的基于文件的电子信息可以被改变而不留下任何的物理证据。在介质和信息系统之间迁移时,基于文件的电子信息也易受到意外的损坏。因此,确保长期基于文件的电子信息真实性的组织宜制定合适的政策、实践和控制技术。常用的控制技术措施包括:

- 使用 WORM(如不可重写式)磁性或光学介质;
- 安全的服务器能够阻止直接利用基于文件的电子信息,提供了“只读”利用的连接;
- 循环冗余校验技术用于检查电子传输可靠性,尤其适用于证实基于文件的电子信息自产生之时未改变;
- 标准散列算法(如 SHA-1)可用一种算法能将基于文件的电子信息压缩成固定长度的位串,可有效地成为基于文件的电子信息的独特“指纹”,以证实它未被更改。

6 长期保存策略的要素

6.1 概述

保存准确、可靠、真实的基于文件的电子信息意味着以下几点:

- 能够通过计算机读取并正确解释;
- 能够以人们理解的格式呈现;
- 具有逻辑和物理结构、实质内容和在信息产生、接收时明显的背景信息。

电子文件存储介质耐久性的局限性和不可避免的技术过时将迫使保存机构为长期保存和利用真实的、可处理的基于文件的电子信息作出关键性选择。面对介质耐久性和技术缺陷的挑战,保存机构有必要采取不同的保存策略和使用不同的工具,这些策略和工具在概念上可划分为以下三项措施,共同构成长期保存策略的基础。

- a) 保存机构宜采取介质更新来确保介质的耐久性;
- b) 如果有自动化工具,将基于文件的电子信息从一个技术平台迁移到另一个技术平台是一种应对技术过时的有效方法;
- c) 当无自动迁移工具时,数字信息和影像被存储在过时系统中时,则需要更有效的措施,在现今的技术环境下,可采用模仿过时的信息系统的方法。尽管这种方法已经有人倡导过,但基于长期利用真实基于文件的电子信息的目的,它面临着操作上的限制。因此,仿真技术在本技术报告中没有进一步讨论。

6.2 介质更新

6.2.1 概述

有限的介质耐久性和技术过时说明介质定期更新是不可避免的,介质更新也是保持原始比特流的“活性”来确保真实和可处理的基于文件的电子信息长期保存的基本需要。介质更新要求基于文件的电子信息被重新格式化或拷贝,6.2.2和6.2.3中有详细介绍。

6.2.2 重新格式化基于文件的电子信息

6.2.2.1 概述

当基于文件的电子信息被重新格式化时,因为转移到不同的物理载体上,它经历了比特流的改变(如从一个18磁道的介质上转移到36磁道的介质上)或者字符代码变更(如从7位ASCII码转到8位ASCII码),但其物理形式或实际内容没有改变。重新格式化不依赖于产生基于文件的电子信息的应用软件。

6.2.2.2 重新格式化的原因

在如下三种情况时,保存机构宜考虑重新格式化基于文件的电子信息:

- a) 迁移时重新格式化:基于文件的电子信息被移交到保存机构时,宜格式化,形成标准编码形式

和存储到标准介质上。

- b) 升级时重新格式化:当保存机构进行设备升级或用新存储设备取代旧设备时,格式化十分必要。
- c) 定期重新格式化:定期格式化宜与使用介质的预期寿命、设备的预期寿命和介质运行的驱动程序的寿命相一致。

6.2.2.3 用于重新格式化的存储介质

保存机构宜慎重考虑存储介质的选择,尤其当重新格式化基于文件的电子信息时。机构宜在磁技术和光技术中进行广泛选择。值得考虑的因素有:

- 存储容量大;
- 数据传送率高;
- 最低预期寿命 20 年;
- 已建立稳固的市场份额;
- 经济上可负担;
- 适合性。

其中存储容量大和数据传送率高是关键因素,因为在介质重新格式化和拷贝时,这两个因素最终决定了基于文件的电子信息的传送时间。当保存机构所存的电子信息以兆或万兆为单位时,这可能成为一个问题。

6.2.2.4 重新格式化和真实性

在重新格式化后,基于文件的电子信息的真实性受到威胁,尤其是存在几种重新格式化情形时。为了提供令人满意的真实性,存储系统和保存机构宜形成书面的质量控制策略,以保证所有重新格式化基于文件的电子信息的准确性。

这项策略的执行程序宜包括重新格式化中所有步骤的全面、完整的使用说明,包括:

- 执行人员的身份;
- 操作日期;
- 数据格式;
- 比较格式化前后生成的循环冗余码和散列摘要值以证实没有改变;
- 若干重新格式化后的基于文件的电子信息与旧格式版本进行视觉比较。

最有效的方法是找出不精确或无法恢复的错误,并将他们保留在后续文件中,宜确定任何无法恢复错误的物理位置(如块、扇区、磁道)。此外,第三方宜检查这些操作,以便确认执行的程序与已建立程序的一致性。这种文件宜标出与基于文件的电子信息的特定连接,并与基于文件的电子信息一样作为元数据得到同样的重视。

6.2.2.5 重新格式化安全

保存机构宜保护基于文件的电子信息在重新格式化的过程中免受更改和丢失。电子存储介质易受人为破坏和灾难性的损失或自然灾害。因此,保存机构宜采取以下几种措施将风险降为最小:

- 宜安装“防火墙”或“单向链接(如”空气间隙“),只允许被授权用户只读访问;
- 电子存储介质宜放置在锁定的安全区域或可控的保险库中;
- 制作存储介质的备份件,并存储在与原件不同的地方;
- 原件与备份件宜使用两种不同的存储介质,以最大程度降低因意想不到的技术过时带来的风险。

6.2.3 拷贝基于文件的电子信息

6.2.3.1 概述

拷贝基于文件的电子信息的目的是通过不改变其结构、内容或背景用同样的格式将其从旧的存储介质迁移到新的存储介质上,以保持其真实性 and 可处理性。当基于文件的电子信息被拷贝到目标存储

介质上时,其内在的比特流保持不变。

6.2.3.2 拷贝原因

保存机构在如下三种情况时需考虑拷贝基于文件的电子信息:

- 迁移时拷贝:当基于文件的电子信息移交到另一个保存机构时宜被拷贝,使用与其移交前相同格式的存储介质;
- 介质有误时拷贝:当每年检查基于文件的电子信息可读性,发现大量样本存在暂时或不可更正的读取错误时,宜拷贝基于文件的电子信息,不必进行介质或设备升级;
- 定期拷贝:当存储介质老化,不必进行介质或设备升级。当现在的版本仍广泛支持并满足机构需要时,宜拷贝基于文件的电子信息。机构宜确定一定时间(如预期介质生命期限的一半)将基于文件的电子信息开始拷贝到可接受的存储介质的新版本中。

6.2.3.3 拷贝真实性

尽管基于文件的电子信息的比特流在拷贝时未被改变,但在此过程中仍存在着被破坏的可能。为了提供令人满意的真实性,保存机构宜有书面的质量控制策略,以保证所有拷贝基于文件的电子信息的准确性。

执行这个策略的程序宜包括拷贝所有步骤的全面、完整的说明材料,包括:

- 执行人员的身份;
- 执行的时间;
- 包括的位数或字节数;
- 比较拷贝前后生成的循环冗余码和散列摘要值以证实没有改变;
- 拷贝的记录信息与旧介质上记录信息进行视觉比较。

最好的方法宜找出不精确或无法恢复的错误,并把处理方法保留在后续文件中。宜确定任何无法恢复错误的物理位置(如块或扇区磁道)。此外,第三方宜检查这些操作,以便确定这些操作与已建立程序一致。最后,已确认存在问题的详细信息宜被清楚地标出来,还要存在与基于文件的信息特定的连接,并作为基于文件的元数据一样得到同样的重视。

6.2.3.4 拷贝安全性

保存机构宜保护电子文件在拷贝过程中免受更改或丢失。电子存储介质易受到人为破坏和灾难性的损坏或自然灾害,保存机构宜采取以下几种措施将风险降为最小:

- 宜安装“防火墙”或单向链接(如“空气间隙”),仅允许被授权用户只读访问;
- 电子存储介质宜放置在加锁的安全区域或可控的保险库中;
- 制作存储介质的备份件,并存储在与原件不同的地方;
- 原件与备份件宜使用两种不同的存储介质,以最大程度降低意想不到的技术过时带来的风险。

6.3 元数据

6.3.1 概述

元数据(关于数据的数据)由信息背景、处理过程和使用时产生的信息构成,它有利于保证真实的基于文件的电子信息确认、检索和保存。

在一些情况下,一些应用软件能自动创建元数据,如文件大小、文件格式、数据、散列摘要值及其他相似属性(如基于文件的电子信息所有权和特性)。在其他情况下,像分类、保存时间、全宗和关键词等这样手工登记的元数据也是十分必要的。这些数据和基于文件的电子信息是可检索的。随着未来各类机构可能会更加广泛的应用企业级内容管理系统,这种情况下,支持保存策略的元数据元素将远比现在得到更多的应用。此外,它们将会自动产生,所以,手工著录将不再需要。相应地,保存机构宜确保元数据捕获和使用工具足够灵活,能够调节大小,以与更多地元数据元素保持一致,便于利用。

6.3.2 可交互的元数据

将来,存在于企业级内容管理系统的元数据将会可交互。元数据的捕获和使用的设计宜考虑将来

可交互的环境中也能被使用,因此,设计机构宜考虑 ISO/TS 23081-1。

6.4 迁移基于文件的电子信息

6.4.1 概述

长期利用策略宜包括提供迁移基于文件的电子信息的措施。

为获取和保护真实的基于文件的电子信息,确保其长期利用,拥有元数据的保存机构面临着四方面挑战:

- a) 可预见的将来,机构和个人在生成和使用基于文件的电子信息时将继续使用各种软件包和存储格式,对保存机构而言,获取可利用的基于文件的电子信息或支持各种软件包和存储格式是非常困难的;
- b) 一些基于文件的电子信息可能依赖于软件,因此,只能在特定的软件环境中使用;
- c) 操作系统和应用软件将不可避免地被更新的、更快的、多功能的系统和软件取代,这就意味着保存机构宜定期将基于文件的电子信息从现在的软件环境转移到新的环境中;
- d) 在缺乏自动迁移工具时,一些基于文件的电子信息只能在保存下来的信息系统中检索。

基于文件的电子信息迁移可以成功地应对上述四项挑战。相应地,保存机构必须保证将真实的基于文件的电子信息从一个应用环境迁移到另一个新的应用环境,而在结构、内容和背景上几乎没有丢失。

6.4.2 软件依赖

长期保存策略宜强调软件依赖的事项。当基于文件的电子信息只能在特定的应用软件中使用时,长期利用基于文件的电子信息是困难的,尤其是软件商不继续支持或提供新的软件版本。在许多情况下,通过损失一些结构可能消除软件的依赖。如当地的文字处理应用中的文本文件信息通过自动嵌入文字处理指令或代码迁移到纯文本中(如简单的 ASCII 文本)。这些指令或代码控制着如字体和脚注等物理属性。

这些操作将减少对软件的依赖。保存机构宜慎重考虑迁移对记录信息真实性的影响。这样的基于文件的电子信息不再被认为是原始的拷贝件,因为它没有复制原始文件信息的结构,甚至可以这样认为,这样产生的基于文件的信息宜被认为是“新文件”,因为其真实性宜通过记录活动而重建,有效性要通过基于文件的信息真实内容没有更改而建立。

将文本电子文件信息迁移到纯文本的另一种方式是将其印刷到纸上或输出到缩微胶片上,这种处理保护了基于文件的电子信息的真实性,尤其适合于基于文件的电子信息的纸张形式,信息被认为在将来可以存储,通过光学字符识别(OCR)使用。

结构型数据库和关系型数据库也能被迁移成平层表格结构以降低对特殊软件的依赖,当相关链接被删除时,每个表格中主要确认和外界的键值宜保留。无论何时,宜建立元数据,以确认这些关系是否是一对一、一对多、多对一或多对多,以便将来重建这种相关性。

6.4.3 软件的升级和安装

对于提供长期利用真实基于文件的电子信息的保存机构而言,软件升级和新软件的安装是难以避免的。长期利用策略为这种必然性提供政策和程序。

当软件升级(如从版本 1 到版本 2)时,开发商在新旧软件之间提供向上兼容性。基于文件的电子信息宜伴随着物理属性,实际内容和信息背景迁移到新的环境中。

当新软件代替现在的软件,或作为独立的应用系统和或一般信息系统升级的一部分,基于文件的信息宜利用旧系统的导出特点和新系统的导入特点进行迁移。此外,将导出/导入路径设计成特殊的专用格式,使一些环境支持迁移(如从一种类型文字处理程序到另一种类型)。

6.4.4 迁移到标准格式

保存机构宜考虑将产生或接收中使用的基于文件的电子信息的大量格式迁移到少量的标准格式上。标准化格式是统一的格式,它们被广泛使用,包括大部分类型的基于文件的电子信息。应避免个性

化的文件格式。值得考虑的中立技术格式是 PDF/A-1, XML, TIFF 和 JPEG。

6.4.5 迁移过时信息系统的基于文件的电子信息

6.4.5.1 概述

当在旧的或包含的过时系统与目标信息系统间不具有向上兼容性,也不存在出口/入口闸道器时,为保证基于文件的电子信息的真实性、可处理性,基于文件的电子信息长期存储策略是迁移。

将来,迁移过时信息系统的基于文件的电子信息也许由于系统的广泛发展而减少,这些系统支持技术中立结构和格式。然而,同时,保存机构为履行其责任,将不得不迁移过时信息系统中的基于文件的电子信息。

在重复的迁移中,由于新旧几代系统之间存在基本的不可兼容性,信息的丢失是不可避免的。因此,保存机构与其试图努力做到信息没有任何丢失,不如考虑制定迁移策略和质量控制程序,以减少迁移过程中的信息损失。重要的是记录迁移时信息丢失情况和质量控制结果。只要有可能,这种记录文件宜与介质一起保存。

6.4.5.2 迁移步骤

6.4.5.2.1 概述

保存机构宜按十步递增方法完成迁移。随着每次迁移环境较大的改变,以下提出的十项步骤不宜所有环境的具体迁移计划。

6.4.5.2.2 分析过时的信息系统(第1部分)

保存机构宜分析过时的信息系统以了解其功能和其中包含的基于文件的信息。包括:

- 各种功能的依据;
- 如何捕获元数据以及元数据与基于文件的信息的关系;
- 基于文件的信息间的关系。

此阶段产生的信息产品宜有详细说明,以便于将“下一阶段工程”的功能、元数据和基于文件的信息应用与新系统。

6.4.5.2.3 分解原有信息系统结构(第2部分)

保存机构宜分解过时信息系统结构,使其界面、应用和数据库服务作为其不同的组成部分——这种想法,不是所有信息系统可以做到,正如下文所述。

- 若系统和用户界面、应用模块、数据库服务和数据库本身是分离的、独立的部分,过时系统可以分解。
- 若界面和数据库是独立的,但应用系统和数据库服务形成一个模块,那么过时系统处于半分解状态。
- 若界面、应用系统和数据库服务被连在一个模块中,那过时系统就是不可分解的。

在任何情况下,为准备迁移,要摆脱对系统结构的外部依赖。

6.4.5.2.4 设计目标界面(第3部分)

目标界面宜提供与原有界面的联系。

6.4.5.2.5 设计目标应用程序(第4部分)

目标应用程序宜提供与原有应用程序的联系。

6.4.5.2.6 设计目标数据库(第5部分)

目标数据库宜提供与原有数据库的联系。

6.4.5.2.7 安装和全面测试目标环境(第6部分)

宜确定、选择、安装、全面测试具备合适安装工具的开放的目标环境。

6.4.5.2.8 产生和安装必要的闸道器(第7部分)

为确保将过时系统的功能一致和准确地复制到目标系统,迁移基于文件的电子信息,宜设计、产生和安装闸道器。闸道器通常有两个作用:一是隔绝选定的部分免受其他变化的影响;二是在中间组成成

分间充当需求和数据的翻译者。用过时的基于文件的电子信息的样本仔细检测闸道器以确保基于文件的信息的一致性和精确性。

6.4.5.2.9 迁移原有的数据库(第8部分)

原有数据库宜被迁移到目标数据库中。

6.4.5.2.10 迁移原有的应用程序(第9部分)

原有应用程序宜被迁移到目标应用程序中。

6.4.5.2.11 迁移原有界面(第10部分)

原有界面宜被迁移到目标界面中。原有界面(如基于字符的菜单和屏幕)可能被图形界面取代。

7 制定长期保存策略

7.1 长期保存政策

为了长期保存基于文件的电子信息的真实性和可处理性,基于文件的电子信息的保存机构宜制定有关详细政策和程序,强调基于文件的电子信息的真实性是其具有凭证作用的关键,确保在处理过程中的统一和一致。此外,它将有助于保存机构应对在法律诉讼中常出现的真实性问题。

最实用的长期保存政策宜包括以下要素:

- 说明为长期保存真实的、可处理的基于文件的电子信息,保存机构的目标、其他保存目标的确认和保存机构的责任;
- 描述保存机构保存基于文件的电子信息的种类,如合法性或物理属性;
- 描述保存机构坚持执行的基于文件的电子信息管理的最佳实践;
- 迁移活动的环境、迁移的方法和原理的确认;
- 对使用审查种类的说明;
- 说明保存机构人员的作用和外界的责任。

7.2 质量控制

按照已建立的规则和程序保存的基于文件的电子信息,一般被认为更真实,在法律程序中更具有可信度。因此,值得信赖的第三方——保存机构,在各种行动中宜实施政策和最佳实践。

有关基于文件的电子信息如何被管理的证据在法律程序中相当重要,因此,宜与移交到保存机构的基于文件的电子信息一样仔细保存。保存证据宜包括所有相关的政策和程序、迁移过程中数据丢失的文件和定期质量检测的结果,这样做是为了确保与政策和程序的一致。

7.3 安全

7.3.1 概述

为提供长期可利用的、真实的、可处理的基于文件的电子信息,保存机构的核心任务是保护电子文件免受更改、改变或丢失。因此,保存机构宜采取6.2.2.3中介绍的方法以履行此责任。

7.3.2 应用/软件访问控制

保存机构宜使用以下的自动程序控制更改和(或)删除电子信息。

任何被更改/删除的基于文件的电子信息宜通过应用程序/软件自动记录,同时记录更改/删除基于文件的电子信息的人员姓名和原因。

为了阻止未经授权的利用,基于文件的电子信息除了通过软件/应用程序被全面检测和证实其有效外,不宜被利用。只有通过经全面检测和证实有效的软件/应用程序才能够访问基于文件的电子信息,以防止未经授权的利用。

所有的历史资料和工作记录宜存储在不可更改的一次写入的介质上,避免任何更改/删除。

7.3.3 物理访问控制

保存机构可以使用以下方法控制对基于文件的信息系统的物理利用:

- 只有授权的个人被允许利用存储设备;

- 出入工作日志宜登记日期、时间、进入安全存储处的人员身份；
- 当授权人员移动存储介质时，宜进行记录。记录宜包括日期、时间、对移动的原因的简要解释；
- 有保存责任的资深人员宜定期审查工作日志以检查人员是否遵守规则。工作日志宜作为本机
构遵守工作规则的证据被保存下来。

7.3.4 防止损失

保存机构宜使用以下方法来保护基于文件的电子信息，以免丢失：

- 保存设备宜置于自然灾害最小的地方，如水灾、火灾、地震或陨石袭击；
- 保存设备宜包括火灾检测和控制系統；
- 建立全面的灾害恢复策略，其中包括基于文件的电子信息的种类，优先考虑存储介质的抢救和恢复；
- 使用磁性介质的保存机构宜将其存储设备远离未屏蔽的大功率电动机（如空调系统）、发电机、变压器和高压线。

7.3.5 安全政策

保存机构宜形成书面的安全程序，包括：

- 将基于文件的电子信息迁移到保存机构的过程中所使用的安全措施；
- 利用控制程序和对程序的监控；
- 存储设备的位置，最大程度降低因自然灾害引起的信息丢失；
- 灾害恢复计划；
- 遵守存储介质公认的使用标准；
- 提供另外一套存储设备，用于备份存储介质，进行灾害恢复。

7.4 环境控制和监测

电子存储介质的相对脆弱性使它的长久性和可读性面临危险。因此，保存机构宜实施环境监控。措施如下：

- 提供合适的环境温度和相对湿度范围，参照制定的标准或权威部门的研究结果；
- 空气过滤系统的设计宜考虑清除环境空气中灰尘颗粒和气体污染；
- 在保存环境中禁止吃食品、喝饮料和吸烟；
- 每年对基于文件的电子信息抽样读取，确定已丢失的或将丢失的信息。

附录 A
(资料性附录)

一些国家电子文件项目及其他选择性的出版物

A.1 概述

此附录的信息仅供参考。它并不是完整详细的说明,读者宜注意与自己权限相关的特定来源指南的确认。

条款 A.2 包括与 ISO/TR 18492 中采用的国家标准相关的国家电子文件保护项目和与这份技术报告相关的政府出版物,款 A.3 包括其他相关出版物。条款 A.4 为进一步阅读资料。

加原文:中英文对照。

A.2 相关的各国家电子文件保护项目

【1】 澳大利亚,国家档案绿皮书:数字文件的保存方法(1999)

Australia National Archives Green Paper: An Approach to the Preservation of Digital Records (1999)

【2】 荷兰,数字资源保存检测平台:内容、背景和现状 荷兰国家档案馆,2001

The Netherland Digital Preservation Testbed. Migration: Context and Current Status National Archives of The Netherlands. 2001

【3】 法国,法国档案的方向 电子和手工管理文件 法国,巴黎,2002

France Direction des Archives de France. Les archives électroniques. Manuel pratique. La Documentation française. Paris. 2002

【4】 南非,政府体系电子文件管理指南(2000)

South Africa. Guide to the Management of Electronic Records in Governmental Bodies (2000)

【5】 英国,英国国家档案馆,电子文件管理、评价和保存指南。第二册,程序 第五章“保护电子文件”

United Kingdom. National Archives of the United Kingdom. Guidelines for the Management, Appraisal and Preservation of electronic Records. Vlo. II, Procedures. Ch 5 “Preserving Electronic Records”

【6】 美国,国家档案与文件管理署,联邦规则第 36 条,第十二章,B 章,1234 部分,“电子文件管理”

United States. National Archives and Records Administration. 36 Code of Federal Regulations Chapter XII, Subchapter B, Part 1234. “Electronic Records Management.”

A.3 被选出版物

【1】 艾姆巴切, B. I. ed. 电子文件三十年, 斯卡克洛出版社, 2003

AMBACHER, B. I. ed. Thirty Years of Electronic Records, Scarecrow Press. 2003

【2】 比尔曼, D. 电子文件保存中的真实性与虚拟性, 数字图书馆杂志, 5(4), 1999. 4

BEARMAN, D. Reality and chimeras in the preservation of electronic records. D-Lib Magazine. 5 (4). April 1999

【3】 道尔, C. M. 真实电子文件: 长期利用的策略, 克哈森协会, 1999

DOLLAR, C. M. Authentic Electronic Records: Strategies for Long-Term Access, Cohasset Asso-

ciates,1999

【4】 道尔, C. M 为数字文件的长期利用选择存储介质 信息管理杂志, 第 33 册, 第 3 号(1999. 7)
DOLLAR, C. M. Selecting Storage Media for Long-Term Access to Digital Records, in The Information Management Journal, Vol. 33 No. 3(July 1999)

【5】 多德雷赫特, L, 易斯德, T 与迈克奈, H. 电子文件完整性的保存, 杜兰特: 克鲁沃学术出版机构, 2002

DURANTI, L., EASTWOOD, T. and MACNEIL H. Preservation of the Integrity of Electronic Records, Oordrecht: Kluwer Academic Publishers Group, 2002

【6】 埃文斯, J 与林德伯格, L. 描述和分析元数据保存文件的能力。
EVANS, J. and LINDVERG, L. Describing and analyzing the recordkeeping capabilities of metadata sets

【7】 哈德斯, M. 数字保存: 数字图书馆的定时炸弹: 计算机与人类, 31(3), 1997, pp. 189-202
HEDSTROM, M. Digital preservation: a time bomb for digital libraries, Computers and the Humanities, 31(3), 1997, pp. 189-202

【8】 亨特, G. S. 保存数字信息, 《如何操作》手册, 纽约: 尼德-舒曼出版商, 2000
HUNTER, G. S. Preserving Digital Information, A How-TO-DO-It Manual, New York: Neal-Schuman Publishers. 2000

【9】 拉瓦锡, B. F. 技术观察报告。开放档案信息系统参考模式: 介绍指南, 可登陆 http://www.dpconline.org/docs/lavoie_OAIS.pdf

LAVOIE, B. F. Technology Watch Report. The Open Archival Information System Reference Model; Introductory Guide, Available electronically at http://www.dpconline.org/docs/lavoie_OAIS.pdf

【10】 李永宏, 斯莱德瑞, O. R, 夏唐与迈克克莱瑞, V. 数字保存中的艺术和方法, 国家标准和技术机构研究杂志。(2002. 1-2)

LEE, KYONG-HO, SLATTELY, O. R. Lu, XIA TANG and MCCRARY, V. The State of the Art and Practice in Digital Preservation. Journal of Research of the National Institute and Technology, (Jan-Feb 2002)

【11】 瑟贝格, J. 避免技术陷阱: 找到数字保存的可行技术基础: 图书与信息资源委员会报告, 华盛顿: 图书与信息资源委员会, 1999. 1

ROTHENBERG, J. Avoiding Technological Quicksand. Finding a Viable Technical Foundation for Digital Preservation: A Report to the Council on Library and Information Resources, Washington, DC: Council on Library and Information Resources, January 1999

【12】 帕尔, R. F 与爱森伯格, J(eds), 国家档案与文件管理署建立电子文件档案馆: 初期的发展建议, 华盛顿 DC 计算机科学与通讯 国家科学 2003

SPROULL, R. f. and EISENBERG, J. (eds), Building an Electronic Records Archive at the National Archives and Records Administration: Recommendations for Initial Development, DC: Computer Science and Telecommunications Board, National Academies of Science, 2003

【13】 伯多维, K. 近年数字保存的技术发展的回顾。数字保护: 国际观点。图书馆与信息资源委员会, 2002. 4

THIBODEAU, K. Overview of Technological Approaches to Digital Preservation in the Coming Years, in The State of Digital Preservation: An International Perspective.

GB/Z 23282—2009

Conference Proceedings. Washington, CD: Council on Library and Information Resources,

April 2002

【14】 胡 博加特, J. 介质稳定性研究, 国家介质实验室, 1996

VAN BOGART, J. 1996 Media Stability Studies, National Media Laboratory, 1996

A.4 进一步阅读资料

【1】 ISO 11799, 信息与文件 档案馆与图书馆材料存储要求

ISO 11799, Information and documentation—Document storage requirements for archives and library materials

【2】 ISO 14721:2003, 空间数字与信息迁移系统 开放的档案信息系统 参考模式

ISO 14721:2003, Space data and information transfer system—Open archival information system—Reference model

【3】 ISO 15801:2004, 电子影像 信息电子化存储 真实性与可靠性建议

ISO 15801: 2004, Electronic image—Information stored electronically—Recommendations for trustworthiness and reliability

【4】 ISO/IEC 17799, 信息技术 安全技术 信息安全管理实用规范

ISO/IEC 17799, Information technology—Security techniques—Code of practice for information security management
