



中华人民共和国档案行业标准

DA/T 47—2009

版式电子文件长期保存格式需求

Format Requirements for Long-term
Preservation of Fixed-layout Electronic Records

2009-12-16 发布

2010-06-01 实施

前 言

本标准由安徽省档案局(馆)提出。

本标准由国家档案局归口。

本标准起草单位：安徽省档案局(馆)、北京北大方正技术研究院有限公司。

本标准主要起草人：李学香、黄玉明、王学武、周建武、胡华平、邓晓文、段丽琼、吴彬松、华婷、陈姗姗、朱霖露。

引 言

版式电子文件是一种用页面化的形式固定呈现文本、图形、图像等信息的文件。出于交换、发布、存档的目的，越来越多的电子文件被制作或转换为版式文件。

本标准基于长期保存和利用的需求确定版式电子文件应具有格式特征，为制订版式电子文件格式标准，开发相应的制作、阅读、打印等软件提供依据。保证电子文件的长期可读、可解析、可理解。

版式电子文件长期保存格式需求

1 范围

本标准明确了版式电子文件长期保存格式的特征。

本标准适用于各级各类档案馆、机关、团体、企业事业单位和其他社会组织遴选归档保存版式电子文件格式，也适用于版式电子文件格式标准的制订与相关软件开发。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本(包括所有的修改单)适用于本文件。

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 18894—2002 电子文件归档与管理规范

ISO 14721:2003 空间数据和信息传输系统 开放档案信息系统 参考模型(Space data and information transfer systems—Open archival information system—Reference model)

ISO 15489-1:2001 信息与文献 文件管理 第1部分：通则(Information and documentation—Records management—Part 1: General)

3 术语和定义

GB/T 18894—2002 和 ISO 15489-1:2001 界定的以及下列术语和定义适用于本文件。

3.1

版式电子文件 fixed-layout electronic records

内容信息被排版在预定义页面中，具有显示一致性的版面固定的电子文件。

3.2

长期 long-term

虑及技术变化(包括支持新载体、新数据格式)和用户群体变化对典藏信息影响的足够长的时间段。这个时间段延伸到无限未来。

[ISO 14721:2003]

3.3

长期保存 long-term preservation

用一种可靠的、科学合理的方式长期(3.2)维护电子文件真实、完整、有效的行为。

3.4

字体 font

具有同一的基本设计的字形图像集合。如：黑斜体。

3.5

字形 glyph

一个可辨认的抽象的图形符号，它不依赖于任何特定的设计。

[GB/T 16964.1—1997, 定义 3.12]

3.6

光栅图像 raster image

由一组按栅格图案排列的像素形成的图像。

3.7

矢量图形 vector graphics

以给定方向与长度数值的线表示的图形。

3.8

稳健 robustness

表现在容错、格式兼容和避免数据缺失等方面的能力。

4 版式电子文件长期保存格式应满足的需求

4.1 格式开放

4.1.1 有公开发表的相应标准和技术规范。

4.1.2 格式标准和技术规范简洁明了, 页面描述语言不应仅为少数厂商认知和掌握。

4.1.3 没有专利和许可的限制。

4.1.4 厂商中立。

4.1.5 有与产品无关的专家组、标准化组织和产业联盟等维护和支持该格式。

4.2 不绑定软硬件

4.2.1 被多种操作系统和硬件平台支持。

4.2.2 文件的阅读不依赖于特定的阅读软件。

4.2.3 使用与设备无关的颜色规范实现准确打印和再现。

4.3 文件自包含

4.3.1 文件呈现的全部信息完全自包含。

4.3.2 文件中必须包括全部字体的字形描述信息或嵌入字体程序信息。

4.3.3 文件中还应包括光栅图像、矢量图形、颜色信息等其他需要呈现的信息。

4.3.4 文件的呈现不依赖于外部对象。

4.3.5 当文件或文件中的部分信息由模拟源转换而来时, 有编码保存其重要属性的能力。

4.4 格式自描述

4.4.1 设置规范的元数据集(可以与国际、国内相关标准建立映射), 以文本方式(通常为 XML)内嵌于文件中, 用于描述文件和对象的属性特征, 并易于提取和检索。

4.4.2 文件中应声明遵循的格式标准和版本。

4.4.3 文件中允许封装用户自定义的元数据。

4.4.4 文件中的字符对象应声明编码标准。

4.4.5 文件有必要且充分的结构信息和语义信息, 用于解析数字对象。

4.5 显示一致性

4.5.1 固定呈现文件页面、章节、段落、字体、图形、图像、色彩等, 呈现效果不因软硬件平台和阅读软件变化而变化。

4.5.2 禁止由条件触发的显示或打印的内容。

4.5.3 禁用音频、视频等多媒体对象。

4.5.4 禁用交互式表单和内部可执行代码。

4.5.5 当使用某些技术将导致显示、打印不一致时应禁止使用, 如透明、隐藏和开放式印前接口(OPI)等。

4.6 持续可解释

4.6.1 允许用户对文件和文件中的对象添加注释，注释应符合 4.5 规定。

4.6.2 不允许设置口令保护。

4.6.3 禁止对文件或文件中的对象实施加密。

4.6.4 文件中运用的算法应该是公开的算法。

4.6.5 文件中引用的字体和运用的算法不应是知识产权保护对象。

4.7 稳健

4.7.1 文件格式稳定，版本升级更新有序，前后兼容。

4.7.2 设置有效的容错机制。

4.7.3 不采用有损压缩、子采样、缩减采样或者其他任何有可能改变文件内容或降低原始数据质量的操作。

4.8 可转换

4.8.1 支持其他格式与版式电子文件长期保存格式相互转换。

4.8.2 支持过时的版式电子文件长期保存格式转换为新的版式电子文件长期保存格式。

4.9 利于存储

4.9.1 格式紧凑，数据结构合理，数据占用字节数少。

4.9.2 支持在一个文件中容纳大量对象和数据。

4.9.3 具有聚合能力，可以把几个相同格式的文件聚合到一个文件中。

4.10 支持技术认证机制

4.10.1 支持数字签名。

4.10.2 使用的数字签名被多种操作系统和硬件平台支持。

4.10.3 数字签名的全部信息完全自包含，包括签名日期、摘要算法、签名算法、签名值、证书信息等。

4.10.4 数字签名必须视觉呈现。

4.10.5 呈现数字签名的电子印章、手写签名等必须自包含。

4.11 易于利用

4.11.1 支持增值服务，方便提取数据的子集，如分割页面、提取文本串、图像等。

4.11.2 支持全文检索。

4.11.3 文件中包含的文本信息符合自然阅读顺序。

5 测试与评估

依据本标准制订版式电子文件长期保存格式和开发软件产品，必须按相关规定经过测试和评估后，才能确认与本标准的一致性。
