

中华人民共和国国家标准

GB/T 42133—2022

信息技术 OFD 档案应用指南

Information technology—Guideline for OFD using in archive

2022-12-30 发布

2023-07-01 实施

国家市场监督管理总局
国家标准化管理委员会 发布



目 次

前言	I
引言	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	2
5 总则	2
5.1 适用场景	2
5.2 处理原则	2
6 长期保存处理	3
6.1 文件结构	3
6.2 基本结构	3
6.3 页面描述	5
6.4 图形	5
6.5 图像	6
6.6 文字	6
6.7 视频	6
6.8 复合对象	6
6.9 动作	6
6.10 注释	6
6.11 自定义标引	7
6.12 扩展信息	7
6.13 数字签名	7
6.14 版本	7
6.15 附件	8
6.16 加密和解密	8
7 软件规则	8
7.1 生成软件	8
7.2 阅读软件	8
附录 A（规范性） 附件转换规则	9
参考文献	10

前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本文件起草单位：国家档案局档案科学技术研究所、中国电子技术标准化研究院、北京数科网维技术有限责任公司、百望股份有限公司、天津市档案馆、宁波市档案馆、江苏省档案馆、上海信联信息发展股份有限公司、方正国际软件(北京)有限公司、北京冠群信息技术股份有限公司、北京华电园信息技术有限公司、北京书生电子技术有限公司、福昕鲲鹏(北京)信息科技有限公司、航天信息股份有限公司、福建正孚软件有限公司、北京海泰方圆科技股份有限公司、北京国脉信安科技有限公司、南威软件股份有限公司、华云数据控股集团有限公司、北京艾卡威特信息技术有限公司、上海泛微网络科技股份有限公司、北京致远互联软件股份有限公司、友虹(北京)科技有限公司、北京点聚信息技术有限公司。

本文件主要起草人：柴德华、董建、聂曼影、刘丹、陈亚军、晏杰、冯辉、吕艳静、仇伟海、陈吉、金雷、蒋余芬、袁嘉新、周凯、彭革非、徐亮、李平立、王少康、祝成、薛馨枫、董艳会、郝立臣、王雷、方俊、周兆峰、李绯、倪时龙、柳晶、郑志梅、王锦妙、丛培勇、黎泽嘉、杨正道、张建、胡守云、张程、杨静、秦聪、黄岩、陆猛。

引 言

GB/T 33190 规定了适用于版式文档的电子文件交换和存储格式(简称 OFD),该格式针对多种业务场景需求而设计,包含多种技术特性。就电子文件长期保存目标而言,GB/T 33190 中的部分内容与电子文件长期保存的要求并不全部一致,在需要长期保存的版式文档中限制使用 OFD 的部分特性(如外部图文资源、加密等)十分必要,以符合 DA/T 47 的规定。同时,为了便于用户准确识别该类文档的处理软件,对这些软件执行特定功能也给出了规则。

本文件的主要内容按照 GB/T 33190 规定的结构进行组织,为方便对照,第 6 章对首次出现的文档模型概念给出了对应的英文短语。任何对 OFD 有效的特性,只要本文件中没有禁止,就可被应用到需长期保存的版式文档中。



信息技术 OFD 档案应用指南


1 范围

本文件给出了使用 OFD 格式的电子文件在档案领域应用的总则、长期保存处理和软件规则。

本文件适用于包含字符、光栅图像和矢量数据的电子文档的生成、转换及处理,也适用于处理该类文档的软件的设计、开发、应用和检测。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中,注日期的引用文件,仅该日期对应的版本适用于本文件;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 13000 信息技术 通用多八位编码字符集(UCS)
- GB 18030 信息技术 中文编码字符集
- GB/T 18894 电子文件归档与电子档案管理规范
- GB/T 33190 电子文件存储与交换格式 版式文档
- GB/T 35275 信息安全技术 SM2 密码算法加密签名消息语法规范
- GB/T 38540 信息安全技术 安全电子签章密码技术规范
- DA/T 46 文书类电子文件元数据方案
- DA/T 47 版式电子文件长期保存格式需求 
- DA/T 54 照片类电子档案元数据方案

3 术语和定义

GB/T 33190 界定的以及下列术语和定义适用于本文件。

3.1

字形 glyph

一个可识别的抽象图形符号。

[来源:GB/T 16964.1—1997,3.12,有修改]

3.2

子集化字型 subset font

只包括部分字符字形轮廓数据且符合字型文件格式(如 OpenType)的要求的字型文件。

3.3

长期保存 long-term preservation

用一种可靠的、科学合理的方式长期维护电子文件真实、完整、有效的行为。

[来源:DA/T 47—2009,3.3]

3.4

合格阅读软件 qualified reader

能够按照本文件给出的规则读取、显示和打印 OFD 文件的应用程序。

3.5

合格生成软件 qualified writer

能够按照本文件给出的规则写出 OFD 文件的应用程序。

注：合格生成软件既可以是单独的软件，也可以是具有导出 OFD 文件功能的任何信息处理系统。

3.6

去技术化 de-technicalization

移除特定的技术表达内容，使用已知或已公开标准的方法模拟呈现或表示其内容的过程。

3.7

归档 archiving

按照国家规定将具有保存价值的电子文件及其元数据的保管权交给档案部门的过程。

[来源：DA/T 58—2014,3.5]

3.8

电子档案 electronic record; archival electronic record

具有凭证、查考和保存价值并归档保存的电子文件。

[来源：DA/T 58—2014,2.2]

3.9

文档 document

OFD 文件中包含一组页面和相关的设置和属性的一种逻辑组织单元。

注：一个 OFD 文件中可以包含多个文档。

4 缩略语

下列缩略语适用于本文件。

BMP 位图(Bitmap)

CMYK 青、品红、黄、黑(Cyan, Magenta, Yellow, black)

JBIG 联合二值图像专家组 (Joint Bi-level Image Experts Group)

JPEG 联合图像专家组 (Joint Photographic Experts Group)

OFD 开放版式文档(Open Fixed layout Document)

PNG 便携式网络图形(Portable Network Graphics)

RGB 红、绿、蓝(Red, Green, Blue)

TIFF 标签图像文件格式(Tag Image File Format)

XML 可扩展置标语言(Extensible Markup Language)

5 总则

5.1 适用场景

本文件适用于以下场景：

- a) 用于归档的、以图文为主要内容的电子文件；
- b) 以查阅和打印为主要应用目标的部分照片类电子档案(不含气象、测绘等专业图像)；
- c) 档案管理过程中产生的统计数据和管理记录。

5.2 处理原则

需长期保存的电子文件按照以下原则转换或处理为符合 GB/T 33190 规定的文件格式。

- a) 具备 DA/T 47 要求的以下特性:格式开放、不绑定软硬件、文件自包含、格式自描述、显示一致性、持续可解释、稳健、可转换、利于存储、支持技术认证机制、易于利用;
- b) 按 GB/T 18894 的要求,以“件”为单位转换或组织为版式电子文件。

6 长期保存处理

6.1 文件结构

OFD 文件的包内结构见图 1。

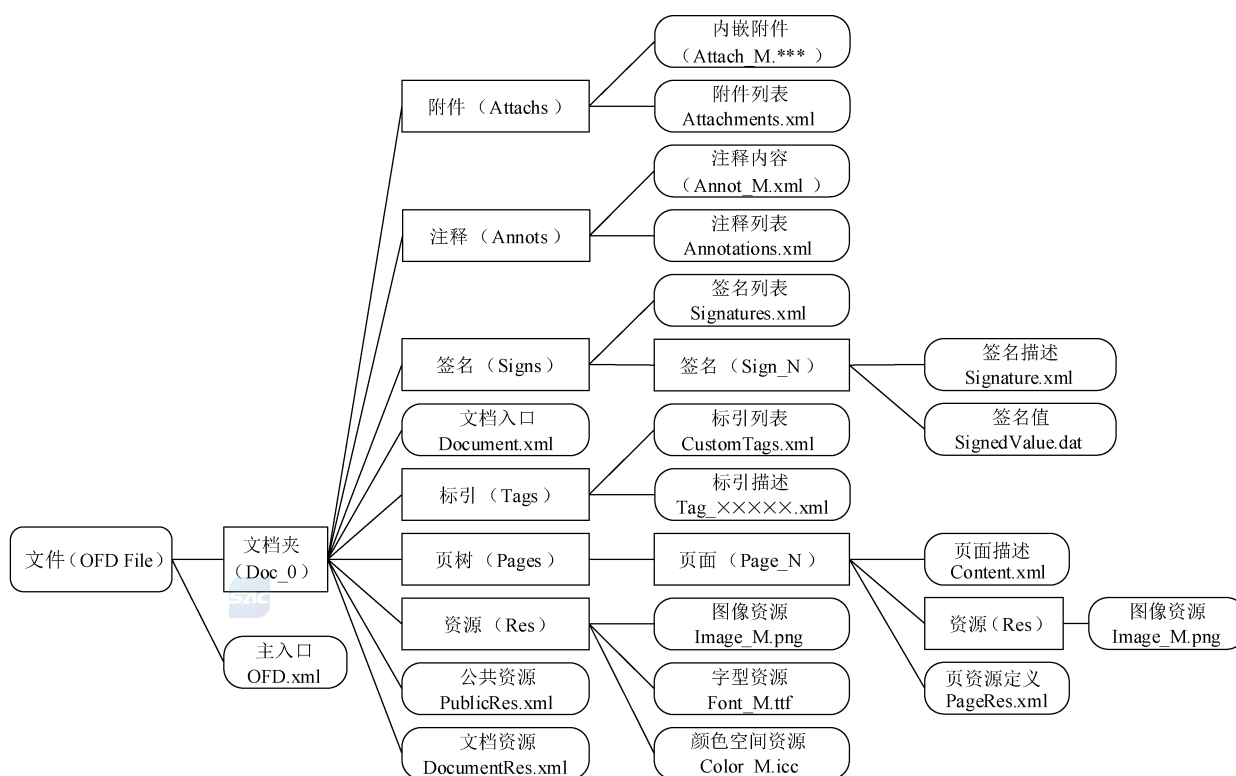


图 1 OFD 文件的包内结构

用于长期保存的 OFD 文件,其内容显示、打印和交互用到的所有数据和关联资源均包含在文件包内。这主要体现在资源文件(PublicRes.xml, DocumentRes.xml 和 Page_N/PageRes.xml)、资源文件夹(Res, Page_N/Res)和注释文件夹(Annots)等处。

用于长期保存的 OFD 文件,还可在上述基本组织之上根据实际需要附加描述该文档背景信息的元数据、符合标准的数字签名或签章、反映源文件结构或内容语义的自定义标引和附件信息等。这主要体现在主入口文件(OFD.xml)中的文档信息(DocInfo)部分或其引出文件、签名文件夹(Signs)、标引文件夹(Tags)和附件文件夹(Attachs)等处。

6.2 基本结构

6.2.1 主入口

对主入口(OFD.xml)文件的处理方式如下。

- a) 主入口文件中根节点的文档类型(DocType)属性取值设置为“OFD-A”。
- b) 文档的元数据以自定义元数据方式存储,元数据名称按 DA/T 46、DA/T 54 的要求确定;当

GB/T 33190 允许以包内独立文件的方式记录元数据时,宜使用该方式。

- c) 归档的 OFD 文件不使用多文档机制。原文件使用该机制的,则在归档前按文档数量拆解为多个 OFD 文件并在每个文档后附加文件列表信息,或者将多个文档内容合并到一个文档中。
- d) 文档中包含多版本(Versions)信息的,保留该信息。
- e) 在 OFD 文件包内去除任何与主入口及其嵌套引出文件无关的文件。

6.2.2 文档根节点

对文档根节点(Document.xml)文件的处理方式如下。

- a) 去除文件中的权限声明(Permissions)。
- b) 去除文件中的视图首选项(VPreferences)。
- c) 去除文件中的类型不是文档内跳转(Goto)的文档动作。
- d) 将文件中不予保留的附件(Attachments)以附件信息页或外部附件方式记录附件的内容摘要值及简要描述。
- e) 去除文件中的扩展信息(Extensions)。当去除的扩展信息影响文件输出效果时,将其影响固化到页面内容中。

6.2.3 页树

对页面描述(Content.xml)文件的处理方式如下:

- a) 将使用最多的页面设置(PageArea)确定为文档的默认设置,并在文档公共数据(CommonData)中描述;
- b) 页面使用的页面设置与文档默认设置相同时,宜在页面描述中省略该节点;
- c) 去除其中类型不是文档内跳转的页面动作;
- d) 对页面内容中名称相同的图层(Layer)重命名,使其可相互区分;
- e) 去除页面内容中的页面块(PageBlock)嵌套,不可避免时,嵌套不宜超过 3 层。

6.2.4 页对象

对页面中内容的处理方式如下:

- a) 图层内容顺序宜按照文档生成时确定,原文的主要内容一般安排在正文层(Body)和背景层(Background);
- b) 通过扫描加工的文档,其扫描图像安排在背景层,识别后的文字内容安排在正文层;
- c) 页面内容流(即 Content.xml 中的内容顺序)宜按照语义顺序组织。

6.2.5 大纲

对大纲(Outline)的处理方式如下:

- a) 去除文档大纲节点(OutlineElem)中类型不为文档内跳转的动作;
- b) 文档大纲节点中的跳转动作跳转至的书签名在书签列表(Bookmarks)中不存在或者跳转的目标页面不存在的,去除该动作。

6.2.6 资源

对资源的处理方式如下。

- a) 颜色空间(ColorSpace)、字型(Font)等在公共资源(PublicRes.xml)文件中定义,栅格图像、音频、视频等多媒体(MultiMedia)、矢量图像(VectorG)、绘制参数(DrawParam)在文档资源(DocumentRes.xml)文件中定义。

- b) 除 GB/T 33190 规定需阅读软件自带的字型和字符范围外, OFD 文件中页面图元使用的字型都嵌入子集化字型数据。“页面图元使用”是指至少出现以下情况之一:
 - 1) 被填充或勾边的文字对象(TextObject)所引用;
 - 2) 被裁剪区中的文字对象(Text)所引用。
- c) 嵌入文档中的矢量字型数据按照 OpenType 规范进行子集化处理:
 - 1) 保留原字型中的 cmap、loca 和 glyf 等字形轮廓相关的表数据;
 - 2) 保留 cvt、fpgm、head、hhea、hmtx、maxp、name、OS/2 和 post 等表数据;
 - 3) 按版式文档字形绘制的需求保留其他表数据。
- d) 字型数据字符集宜与 GB 18030 一致, 并与 GB/T 13000 建立映射关系。
- e) OFD 文件中使用的栅格图像格式限于 BMP、JPEG、PNG、JBIG2、JPEG2000 和 TIFF 六种, 其他格式的栅格图像, 在归档前的处理过程中转换为此六种格式之一。
- f) OFD 文件中使用的栅格图像资源, 不使用其扩展机制加入自定义数据。
- g) 去除 OFD 文件中的音频、视频等资源, 并以外部附件、注释备注(Remark)等方式保留音视频的摘要信息。

6.3 页面描述

6.3.1 颜色

对颜色的处理方式如下:

- a) OFD 文件中可定义默认颜色空间(DefaultCS);
- b) OFD 文件中使用的颜色空间类型限于灰度值(Gray/GRAY)、RGB、CMYK 之一;
- c) OFD 文件中使用的颜色空间宜带有颜色配置文件(Color Profile)。

6.3.2 裁剪区

对裁剪区的处理方式如下:

- a) 裁剪区的实际生效区域包含图元的外接矩形(Boundary)时, 去除该裁剪区;
- b) 裁剪区的实际生效区域面积为 0 时, 去除该裁剪区并将所在对象的可见性(Visible)设置为“false”。

6.3.3 页面对象

对页面对象的处理方式如下:

- a) 除特殊情况外, 图元外接矩形处理为以能完整包含文字对象的可见部分的最小矩形为基础;
- b) 去除页面图元对象的动作, 图元对象的文档内跳转动作处理为链接注释;
- c) 文本对象、图形对象的是否勾边(Stroke)属性取值为“false”时, 宜去除线宽(LineWidth)、线端样式(Cap)、连接样式(Join)、连接点截断值(MiterLimit)、虚线样式(DashPattern、DashOffset)、勾边颜色(StrokeColor)等配合的修饰属性;
- d) 复合对象的是否勾边、是否填充(Fill)、线宽、线端样式、连接样式、连接点截断值、虚线样式属性以及勾边颜色(StrokeColor)、填充颜色(FillColor)宜去除;
- e) 图像对象、视频对象的是否填充属性以及填充颜色宜去除。

6.4 图形

对图形对象(PathObject)的处理方式如下:

- a) 勾边颜色、填充颜色在多个图形对象之间重复的, 宜注册为绘制参数, 由图形对象或其所在图

层引用；

- b) 线宽、线端样式、连接方式、连接点截断值、虚线样式等绘制属性在多个图形对象之间重复时，宜将共用的绘制属性注册为绘制参数，由图形对象或所在图层中引用。

6.5 图像

对图像对象(ImageObject)的处理方式如下：

- a) 在多个页面中共同使用的栅格图像宜在文档资源文件中注册，在一个页面内多个图元对象间共用的栅格图像宜在页面资源(Page_N/PageRes.xml)文件中注册；
- b) 若栅格图像对象中使用了插值绘制(Interpolate)属性，将其值设置为“false”。

6.6 文字

对文字对象(TextObject)的处理方式如下：

- a) 在一行内相邻且字体属性一致的字符宜使用一个文字对象描述，多行字符内容宜使用多个文字对象描述；
- b) 使用了 GB/T 33190 规定需阅读软件自带的字型，但部分字符代码(TextCode)在指定范围之外的，拆分为两个或多个文字对象；
- c) 文字的大小使用字号(Size)标识；
- d) 文字对象仅使用横向缩放变换时，宜使用水平缩放(HScale)标识；
- e) 文字的勾边颜色或填充颜色在多个以上文字对象之间重复的，宜在绘制参数中注册，由文字对象或所在图层引用。

6.7 视频

视频对象按带有视频动作的图像对象处理。



6.8 复合对象

在多个页面中共同使用的矢量图像资源宜在文档资源文件中注册，在一个页面内多个图元对象间共用的矢量图像资源宜在页面资源文件中注册。

6.9 动作

对动作(Action)的处理方式如下：

- a) 图元对象、大纲节点、页面和文档均不保留依赖外部资源的动作；
- b) 去除图元对象、大纲节点、页面和文档中的动作时，宜使用所附着对象的自定义参数机制保留其摘要信息。

6.10 注释

6.10.1 归档前的注释

归档处理过程中保留已存在的注释信息，具体可使用下列方式之一。

- a) 继续作为注释对象存在，但其自定义参数满足下列条件：
 - 1) 去除对注释外观显示有影响的参数，并将其影响固化到注释外观中；
 - 2) 去除注释外观使用的页面块嵌套；
 - 3) 将注释对象的是否只读(ReadOnly)、是否缩放(NoZoom)、是否旋转(NoRotate)属性设置为“true”。

- b) 将该注释对象的外观显示合并到所在页的页面内容中。

6.10.2 归档后的注释

OFD 文件归档后,管理部门在内部管理中需要添加归档章、检验章、页码等内容时,使用注释等方式附加到文档中,并与文档的原有注释对象分文件存储。

6.11 自定义标引

归档前已存在的自定义标引宜保留,且将标引描述或引用的全部内容包含在当前文件内,不依赖于外部内容或资源。

6.12 扩展信息

对扩展信息(Extensions)的处理方式如下:

- a) 去除 OFD 文件中的扩展信息;
- b) 扩展信息的对象是图元、图层、页面,且扩展信息对该对象显示或打印效果有影响的,将其影响固化到对象的外观或页面内容中。

6.13 数字签名

6.13.1 签名数据格式

OFD 文件中数字签名至少包括签名和签章两种类型,其满足以下条件时可不“去技术化”:

- a) 签名数据按照 GB/T 35275 的要求组织,且该结构中的公钥证书由有资质的 CA 机构颁发;
- b) 签章数据按照 GB/T 38540 的要求组织,且该结构中的电子印章已备案,其制章人公钥证书、签章人公钥证书均由有资质的 CA 机构颁发;
- c) CA 机构的证书状态查询服务或吊销证书列表服务的可持续性被档案主管部门认可。

6.13.2 签名的去技术化

签名(Type=Sign)的去技术化处理规则如下。

- a) 被注释对象引用且保留注释对象时,按照如下顺序处理:
 - 1) 在注释对象的自定义参数中增加签名人、签名时间和签署时文档摘要值等信息;
 - 2) 去除注释对象中的签名引用(SignRef)属性。
- b) 被注释对象引用且不保留注释对象时,将注释对象的外观转换为页面内容。
- c) 从数字签名列表(Signatures.xml)文件中去除该签名对应记录,并移除对应的文件夹。

6.13.3 签章的去技术化

签章(Type=Seal)的去技术化处理规则如下。

- a) 对签章外观进行处理,具体可按以下方式之一:
 - 1) 将签章外观转换为注释对象,并在该注释对象的自定义参数中保留签章人名称、证书编号、签章时间、签署时文档摘要值等信息;
 - 2) 将签章外观转换为页面内容,并设置其图像混合模式(BlendMode)为“Darken”。
- b) 从数字签名列表(Signatures.xml)文件中去除该签章对应记录,并移除对应的文件夹。

6.14 版本

归档处理时保留 OFD 文件内包含的多版本信息。

6.15 附件

6.15.1 附件保留

OFD 文件中的下列附件宜保留。

- a) 使用 TXT、XML 等文本格式。
- b) 所描述的内容有公开发布的标准或技术文档说明其组成,且相关技术文档已归档。
- c) 在符合 a)和 b)的基础上包含数字签名且该签名数据按照国家标准组织的。

6.15.2 附件去除

OFD 文件中有附件且不予保留时,按照附件的格式类型分别处理。

- a) 按照附录 A 可转换为版式文档的,转成版式文档并将转换后的内容依次插入到主文档后。
- b) 下列情况宜将附件从 OFD 文件中移出,并以附加附件信息页或外部附件方式在文档内保留移出附件的摘要信息:
 - 1) 附件是音频或视频文件;
 - 2) 附件是超大尺寸图像文件,且不以查阅和输出为主要应用目标;
 - 3) 附件是数据类电子表格文件;
 - 4) 附件是网页或演示文稿文件,含有较多动画、三维效果且不宜丢失。

6.16 加密和解密

用于长期保存的 OFD 文件不使用任何加密选项,包含相关选项的,在归档前解密为明文。

7 软件规则

7.1 生成软件

合格生成软件具有将多种文件和数据转换为或生成为 OFD 文件以及将 OFD 文件处理为适于长期保存的功能。

合格生成软件将 OFD 文件处理为适于长期保存时,按照本文件给出的原则和方式进行。

7.2 阅读软件

合格阅读软件具有解析并完整输出 OFD 文件以及按长期保存处理规则进行文件格式检查的功能,进行格式检查后,对其中的未处理项以对话框、日志等方式加以提示。

合格阅读软件在打开文档类型属性取值为“OFD-A”的 OFD 文件时,禁用插入页面、调整页面等文档编辑功能,并以提示栏、页面水印等方式进行提示。

附 录 A
(规范性)
附件转换规则

文档中的附件转换规则按表 A.1。

表 A.1 文档附件的转换规则

序号	类型	子类型	转换前的文件格式	转换后的文件格式及说明
1	图文类	图像文件 *	bmp、jpeg、png、jbig2、jpeg2000 和 tiff 等	按本文件处理后的 ofd 带 * 号的还应根据内容和应用 目标综合判断
2		文本文件	txt 等	
3		文字处理文件	doc、docx、wps、rtf、uof、uot 等	
4		电子表格文件 *	xls、xlsx、uos、et、wpt 等	
5		演示文稿文件 *	ppt、pptx、wpp、uop 等	
6		版式文件	ofd、pdf、xps、ps、ceb、sep、aip 等	
7		网页文件 *	html、htm、asp、jsp 等	
8	音频类	音频文件	wav、mp3 等	avs
9	视频类	视频文件	avi、mpg、mp4、flv、mxf 等	avs
10	其他类	计算机辅助设计类	dwg、dxf、igs、dat 等	若文件的主要用途是查阅和输出,可转换为版式文件
11		数据库类	mdf、mdb、db、dbf、wdb 等	可按照纸质或应用表单样式,转换为版式文件;使用数据库文件保存的照片档案可参照图像文件处理
12		电子邮件	eml 等	可按照网页文件转换或按应用表单合成为版式文件
13		社交媒体类文件	html、txt、png 等	可参照文本、图像或电子表格文件处理方式转换为版式文件
<p>注 1: 表中文件格式指的文件的后缀名。</p> <p>注 2: avs 文件格式见 GB/T 20090(所有部分)。</p>				

参 考 文 献

- [1] GB/T 16964.1—1997 信息技术 字型信息交换 第1部分:体系结构
 - [2] GB/T 20090(所有部分) 信息技术 先进音视频编码
 - [3] DA/T 1 档案工作基本术语
 - [4] DA/T 18 档案著录规则
 - [5] DA/T 47—2009 版式电子文件长期保存格式需求
 - [6] DA/T 58—2014 电子档案管理基本术语
 - [7] DA/T 80 政府网站网页归档指南
 - [8] DA/T 85 政务服务事项电子文件归档规范
 - [9] ISO/IEC 14496-22 Information technology—Coding of audio-visual objects—Part 22:
Open Font Format
-

