

长久保存格式梳理

第二组 王墨竹 张梦雪 路 莹
黄碧瑶 王小丹 廖婷婷

1.电子文件长久保存的重要意义

2004年联合国教科文组织发布《保存数字遗产》宪章，该宪章开宗明义地指出：“任何形式的遗产的消失都是全人类遗产的损失。数字化遗产是共同遗产。”在这个信息化社会，数字资源日益增多，包括文字、数据库、静止的和活动的图像、声音和图表、软件和网页等等，他们很多存在时间不长，在互联网的动态更新中很快就会消失，需要有意地维护与管理才能长久保存下来。电子文件无疑是数字遗产的重要组成部分，所以电子文件长久保存不仅仅是档案领域所关注的重大事情，也是全人类的责任。如何管理好电子文件，以加强质量管理、改善管理的有效性和提高管的效率，选择和长久保存那些值得为子孙后代留存的电子文件，帮助公众、学者和政府官员发现且获得我们国家历史上任何时期他们感兴趣的电子文件，显然非常重要。档案是现在留给未来的财富，只有我们现在能够完整保存好这些有价值的信息，才能让子孙后代有丰富的精神财富和文化遗产。

在日常的工作、生活中，不论是个人还是企业机构都有长久保存某些信息的需求，然而，尽管大多数人都了解电子文件长期保存的重要性，却没有要保存数字信息的意识和紧迫感。其原因一是乐观的，认为未来一定会有新的软件和读取设备发明，不会存在数字信息无法解读的问题；二是悲观的，认为电子文件易受病毒攻击而损坏、丢失，并且存取设备换代快，无法长期保存。这两种想法都是极端的，我们既需要对电子文件的保存技术有足够的信心，又不能盲目乐观，要认识到电子文件长期保存的难点，采取适当的防范措施。

电子文件长期保存是一项艰巨而复杂的系统工程，涉及诸多主体、环节、层次和领域，文件或档案管理部门单凭一己之力是难以胜任的，需要在不同层次、范围内加强协调与合作。在国际上，对于数字信息长期保存采取合作策略已经是一个基本共识，至于合作方式、合作主体、合作机制等问题各国都在研究，在探讨与实践当中不断前进。

2.长久保存格式的梳理

2.1 电子文件常见格式

按电子文件的格式划分，有数字文本文件、数字影像文件、数字图形文件、数字语音文件、数字视频文件。

文本文件常见格式有 XML、TXT、UOF、RTF、WPS、PDF/A、CEB、SEP；

图像文件常见格式有 TIFF、JPEG2000、JPEG、GIF、PNG、DjVu；

音频文件常见格式有 WAV、MP3、WMA、OGG Vorbis 、FLAC； 视频文件常见格式有 AVI、MOV、MPEG 等。

2.2 长期保存格式选择存在的问题

选择专用媒体或是专用存储格式将来会导致数字存储问题，例如媒体格式。选择专用媒体格式存储电子文件，必须考虑将来该格式的过时问题。再例如文件格式。为了确保电子文件的长期存取，如果采用专用的文件格式，会由此增添更大的复杂性。专用软件的应用常常升级至最新版本，然而软件的功能却没有像版本变化那么明显。随着时间的过去，文件格式累积的变化更为显著，甚至危及格式逆兼容。对遗产软件的维护显得很急迫，但是存在的问题颇多。就像应用软件一样，操作系统也得定期升级，随着系统基本体系结构的发展，将来不再支持遗产软件，不对遗产软件支持就意味着产生新的安全问题的风险也在增加。

2.3 数字文本文件保存格式探析

在数字文本文件长久保存格式梳理中，我们将具体格式分成 4 类：

《电子文件归档与管理规范》推荐文本文件格式——XML、RTF、TXT；

国内数字图书馆（电子期刊网络）主流格式——CAJ、SEP；

办公软件主流格式——CEB、WPS、PDF；

标准化主流格式——UOF、OOXML；

2.3.1 《电子文件归档与管理规范》推荐文本文件格式

中华人民共和国国家标准《电子文件归档与管理规范》（GB/T18894-2002）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。其中特别要求到，要能够对档案数据采用通用格式，或能够转换成通用格式，以便实现档案信息资源的共享，其中推荐的通用文件格式为 XML 文档和 RTF、TXT 格式。

（1）XML

格式说明：

Extensible Markup Language，即可扩展标记语言，是 W3C 组织（万维网联盟）于 1998 年 2 月发布的标准。W3C 组织制定 XML 标准的初衷是定义一种互联网上交换数据的标准；从其本质上看，XML 也是一种标记语言，基本上是 SGML 的一个子集。

优点与缺点：

首先，XML 作为一项国际标准，可以方便不同系统之间进行信息资源的交流，便于数字化档案的迁移。同时，XML 的内容与外观设计是完全分开的，外观变动时，XML 文件本身不受影响。在有一定需要进行信息搜索时，只要对 XML 文件进行搜索，就可以得到自己想要的有用信息。XML 具有方便的信息搜索，这一点对内容庞大、复杂的档案全文数据库来说是很重要的。但同时，XML 格式制定初衷就是应用于互联网，故 XML 文件更适合于网络应用，早期的基于 C/S 模式的档案管理系统也许不能很好的对其进行管理。

(2) RTF

格式说明:

Rich Text Format, 即丰富文本格式, 又叫做多(信息)文本格式, 与 DOC 文件格式有些类似。它不仅可包含传统的文字及其格式信息, 还可包含图像、图形等多种媒体信息, 是一种非常流行的文件格式, 很多文字编辑器都支持它。

优点与缺点:

从本质上讲, RTF 格式是一个很好的文件格式转换工具, 用于在不同应用程序之间进行格式化文本文档的传送。但 RTF 格式有一个显著缺点是有可能进行扩展。例如微软的 WORD 就有自己的 RTF 命令, 而这些命令可能在其他的 RTF 编辑器中不能被认识。

(3) TXT

格式说明:

TXT, 即纯文本格式, 是微软在操作系统上附带的一种文本格式, 也是最通用的文本文件格式。它主要用于保存文本信息, 即为文字信息, 现在的操作系统大多使用记事本等程序保存, 大多数软件可以查看, 如记事本, 浏览器等。

优点与缺点:

TXT 文件体积相比其他众多格式较小, 阅读不受限制, 几乎所有的文字处理软件都能识别, 使用也非常方便, 如 Windows 自带“附件”中的“记事本”程序即可打开阅读。然而, TXT 格式缺点之一是不能插入图片, 显然会在极大程度上影响档案的完整性与原始性。此外, 虽然自 Windows98 后, TXT 不再受限制于自身仅有 64KB 的容量, 如果仅做为内部存档, TXT 格式还勉强可以胜任, 但由于实际工作中很少使用 TXT 文件, 与外界进行交换时也会受到较大限制。

(4) 三种格式的共通性与缺点

总体而言, 三种格式都存在格式通用性好, 可移植性强等优点, 比较适合作为数字化档案保存与利用的格式; 但同时, 这三种格式也存在着共同的缺陷, 即易于修改而不能保证档案的原始性, 或在格式转换过程中尽管对原始信息内容进行了保留, 却对原始信息格式进行了调整。例如在将 WORD 格式转化为 TXT 格式时, 就会对原有的 WORD 格式做出很大改变。电子文件长久保存的核心点就在于保证其真实性、原始性、可靠性, 方能发挥其电子凭证的作用; 而三种格式所存在的真实性与原始性的不确定性, 也在很大程度上影响了其作为长久保存的电子文件格式的可利用性。

2.3.2 国内数字图书馆(电子期刊网络)主流格式

(1) CAJ

格式说明:

China Academic Journals, 原为中国学术期刊全文数据库英文缩写, 后为清华大学《中

国学术期刊(光盘版)》电子杂志社在“中国期刊网”的网络出版中,开发的 CAJ 文件格式,主要是为学术期刊上网而开发的。它支持中国期刊网的 CAJ、NH、KDH 和 PDF 格式文件。

优点与缺点:

CAJ 封装文字及插图信息在单一的文件中,完整地保留原来文件的版式信息,且打印效果与原版的效果一致。除此以外,CAJ 还有打印全文功能和机上摘录功能,可以放大和缩小页面,并有在当前页面内的查找字符功能。但同时,CAJ 格式也存在着其较明显的局限性,即更多应用于学术期刊和论文,应用面并不广泛。究其原因,还在于 CAJ 文件在制作过程中的使用格式与字符集和 WORD 并不兼容,这也是导致 CAJ 格式无法很好地与其它格式文件进行资源的交换与信息的传递的重要原因。

(2) SEP

格式说明:

北京书生公司开发的基于二进制 XML 技术的 SEP 文件格式,是博采 XML, PDF, S2 等各种技术之长的自有版权的新一代 DPaper 的通用格式。SEP 技术是书生公司研发的新一代数字纸张技术的核心,基于 JBIG2 的黑白图像压缩和 JPEG2000 的彩色图像压缩使得文件体积大大缩小。文件的扩展名为.SEP。

优点与缺点:

SEP 格式与前文的 CAJ 格式一样,是目前国内数字图书馆应用最多的文件格式之一。文件在经过转换后可以完整保持转换前的版式,嵌入相关的字库和图片,使原始文件中的图片、表格、色彩、补字、艺术字等复杂的版式与内容都能被准确的保留,并保证在不同的计算机环境下版式也不会发生变化。但同时,SEP 文件只能进行浏览、打印、盖章等限定操作,而不能随意编辑与修改,虽然对文档信息的内容进行了有效的保护,但在一定程度上也有所限定。

2.3.3 办公软件主流格式

(1) PDF

格式说明:

Portable Document Format,即可移植文档,又叫做便携式文档格式,由 Adobe 公司开发。PDF 是一种通用文件格式,能够保存任何源文档的所有字体、格式、颜色和图形,无论创建该文档所使用的应用程序和平台。即,无论字体、软件和操作系统,PDF 始终如原文件一样显示。目前已成为全世界电子文档分发的公开的实际标准。

优点与缺点

PDF 格式的优点非常瞩目,首先,它可以很好的保持档案的原貌。PDF 格式能够做到忠实再现原稿的每一个字符,颜色以及图像。不管显示器是何种类型,PDF 文件都能够放大到 800%而丝毫没有损失。究其根本,PDF 同时独立于软件、硬件和创建该文件的操作系统。它是

Post Script 语言图像模型为基础，无论在何种打印机上都可保证精确的、颜色准确的打印效果。其次，PDF 文件比源文件小很多，在 web 上下载文件的同时可以快速地显示页面，而并不会在同时降低网络速度。利用者不需要下载整个 PDF 文件然后再阅读，从这个意义上说，PDF 非常适合网上传输，而这与当前档案信息电子化的工作趋势非常吻合。第三，PDF 可以满足档案工作的需求。对于档案管理者来说，PDF 格式有很强的安全性。因为它能够控制文件的访问权限，可以防止他人复制、改变、打印 PDF 上的文本和图像。对于档案利用者来说，PDF 可以方便地通过浏览器浏览，Acrobat Reader 可以自动打开 PDF 供读者阅读，利用率与便利度都很高。目前，PDF 已成为事实上的工业标准，广泛用于各行业的文档传递，利用已经较为成熟。

PDF 的缺点也同时存在，例如它压缩比较高，相较于其他格式（如 CAJ）而言不利于网上传送；更重要的是，PDF 丧失了其编辑的功能，变为“只读”，想要对它排版十分困难，因为排版信息都已经丢失。即便 Windows2013 花费极大功夫增进其编辑功能，效果仍欠佳。

(2) CEB

格式说明：

Chinese eBook，北大方正开发的完全高保真的中文电子文档格式，目前我国政府机关公文处理中应用广泛。常用作文字处理排版软件，如 WORD、WPS 等，生成的所有结果文件都可以方便地转换成 CEB 格式。

优点与缺点：

首先，CEB 可以较好地保持文件原版样式，能够保留原文件的字符、字体、版式和色彩，包括图片、数学公式、化学公式、表格与文件中的圈注、批注、划线等一切信息。同时，CEB 格式文件一旦生成，便不能随意删改和篡改，因它所具有有数字签名、数据加密功能，可以使档案管理人员在保护档案知识产权的基础上，有权限地为用户提供利用，从而保证了公文正确、完整和保密，符合国家行政机关对公文格式的严格要求，也符合数字化档案的保存与利用需求。除此以外，CEB 还支持汉语、少数民族语言、英语等多种语系，可以完成自动分类、自动摘要、图片检索等操作，很方便用户的使用，更可以极大地减轻档案管理人员的工作负担。最后，CEB 还可以加入多种压缩方法，从而对文字及图像信息进行很好压缩，使文档保持最小，因此数据量小，传输及存贮较为方便等；它支持多种颜色空间和色彩控制，支持复杂的表单设计，有较好的屏幕与打印输出效果。

(3) WPS

格式说明：

Word Process System（又译 Word Processing System），国内著名的办公软件，在 DOS 时代占据了国内办公软件的市场。1988 年到 1995 年的 7 年间，WPS 几乎成了电脑的代名词。但在向 Windows 过渡时，没有把握住机会失去了市场。凭借操作系统的优势，微软 Office 套

件逐步占领了市场。目前，全新的 WPS office 办公软件功能强大，完全可以替代微软的 office 套件，国内的政府机构和企事业单位也有着广泛的应用，成为上至国务院 57 部委、下至全国 31 个省市机关的标准办公平台。

优点与缺点：

WPS 作为一款更推向了技术型中低端客户使用的办公软件，兼容性能极佳，体积小，且适合多种界面切换及“云”办公。同时，它也拥有诸如一系列的加密信息保障，版权保护技术，可扩展的插件机制等。然而，在专业性方面，WPS 缺少 VBA 宏语言的支持，在大文件的处理方面也有自己的漏洞。

2.3.4 标准化主流格式

(1) UOF

格式说明：

Uniform Office document Format，中文译作标文通，全称为中文办公软件文档格式规范。UOF 采用 WC XML Schema 作为描述文字处理、电子表格和演示文档格式信息的语言；同时也定义了针对物理存储格式以及文档标准应用过程中的辅助性规范。此规范作为中文办公软件文档格式和存储格式的定义，支持不同的中文办公软件之间的兼容和文档互换。目前，主要的国内办公软件均已支持“标文通”。

优点与缺点：

作为流式文档标准之一的 UOF 是我国自主知识产权的中文办公文档格式标准，采用 W3CXML Schema 作为文档格式标准定义语言，形成独立、完整、开放及可扩展的文档描述体系结构，方便用户扩展；文档标准体系架构支持模块的可重用性，减少文档描述的冗余，保证文档简洁易用。总之，自 2007 年作为国家推荐性标准正式发布实施以来，以其诸多的优势得到越来越广泛的应用。

(2) OOXML

Office Open XML，是由微软公司为 Office 2007 产品开发的技术规范，现已成为国际文档格式标准，兼容前国际标准开放文档格式和中国文档标准“标文通”（即 UOF），并于 2006 年 12 月成为 ECMA 标准。OOXML 主要是将文档看做一个容器，规定了其中的各个组成部分。除了少数代表图形的二进制数据或内嵌 OLE 对象，大多数部件都采用 XML 文件形式，这些部件相互之间通过关系协同工作。Office open XML 基本上覆盖了办公软件的各个基本功能，如描述文字处理文档的 word procesing ML、描述电子表格的 spread shet ML、描述演示文稿的 presentation ML 等，同时也增加了对矢量图形（d rawing ML 和 VML）、数学公式（math ML）、用户 XML 数据等的支持。特别的是，微软 Office 实现了对 OOXML 的部分支持。

2.4 音频文件保存格式探析

● 音频部分

(1) WAV

格式说明:

WAV 为微软公司 (Microsoft) 开发的一种声音文件格式, 它符合 RIFF(Resource Interchange File Format) 文件规范, 用于保存 Windows 平台的音频信息资源, 被 Windows 平台及其应用程序所广泛支持, 该格式也支持 MSADPCM, CCITT A LAW 等多种压缩运算法, 支持多种音频数字, 取样频率和声道, 标准格式化的 WAV 文件和 CD 格式一样, 也是 44.1K 的取样频率, 16 位量化数字, 因此在声音文件质量和 CD 相差无几。WAV 打开工具是 WINDOWS 的媒体播放器。WAV 是最接近无损的音乐格式, 所以文件大小相对也比较大。

优点:

简单的编/解码 (几乎直接存储来自模/数转换器 (ADC) 的信号)、普遍的认同/支持以及无损耗存储。

缺点:

需要音频存储空间。对于小的存储限制或小带宽应用而言, 这可能是一个重要的问题; WAV 格式的另外一个潜在缺陷是在 32 位 WAV 文件中的 2G 限制, 这种限制已在为 SoundForge 开发的 W64 格式中得到了改善。

(2) MP3

格式说明:

MP3 是一种音频压缩技术, 其全称是动态影像专家压缩标准音频层面 3 (Moving Picture Experts Group Audio Layer III), 简称为 MP3。它被设计用来大幅度地降低音频数据量。利用 MPEG Audio Layer 3 的技术, 将音乐以 1:10 甚至 1:12 的压缩率, 压缩成容量较小的文件, 而对于大多数用户来说重放的音质与最初的不压缩音频相比没有明显的下降。它是在 1991 年由位于德国埃尔朗根的研究组织 Fraunhofer-Gesellschaft 的一组工程师发明和标准化的。用 MP3 形式存储的音乐就叫作 MP3 音乐, 能播放 MP3 音乐的机器就叫作 MP3 播放器。

优点:

它丢弃掉脉冲编码调制 (PCM) 音频数据中对人类听觉不重要的数据 (类似于 JPEG 是一个有损图像压缩), 从而达到了小得多的文件大小。

MP3 音频可以按照不同的位速进行压缩, 提供了在数据大小和声音质量之间进行权衡的一个范围。

MP3 不仅有广泛的用户端软件支持, 也有很多的硬件支持比如便携式媒体播放器 (指 MP3 播放器) DVD 和 CD 播放器。

缺点:

MP3 为了压缩而破坏了音乐的质量;传统 MP3 需要带耳机才有很好的音质,但是对人们的耳膜有所伤害

(3) WMA

格式说明:

WMA(Windows Media Audio),它是微软公司推出的与 MP3 格式齐名的一种新的音频格式。由于 WMA 在压缩比和音质方面都超过了 MP3,更是远胜于 RA(Real Audio),即使在较低的采样频率下也能产生较好的音质。一般使用 Windows Media Audio 编码格式的文件以 WMA 作为扩展名,一些使用 Windows Media Audio 编码格式编码其所有内容的纯音频 ASF 文件也使用 WMA 作为扩展名。

优点:

WMA 7 之后的 WMA 支持证书加密,未经许可(即未获得许可证书),即使是非法拷贝到本地,也是无法收听的。

低比特率的情况下,同文件比 MP3 体积小一倍而音质不变。

微软公司在 WMA 9 大幅改进了其引擎,实际上几乎可以在同文件同音质下比 MP3 体积小 1/3 左右,因此非常适合用于网络串流媒体及行动装置。

缺点:

在高位元率的渲染能力低下,同音源的一个 320kbps 的 MP3 与比较 192kbps 的 WMA 相比,音质和渲染力很容易分出是前者较优。与 MP3 相同,WMA 也是有损数据压缩的档案格式,对于有更高要求的用户来说 WMA 并不是一个适合的格式。(9.0 版本支持无损压缩);WMA 也与 MP3 一样同为有专利版权的档案格式。支持的装置需要购买使用版权。

(3) OGG Vorbis

格式说明:

Ogg 全称应该是 OGG Vorbis 是一种新的音频压缩格式,类似于 MP3 等现有的音乐格式。Vorbis 是这种音频压缩机制的名字,而 Ogg 则是一个计划的名字,该计划意图设计一个完全开放性的多媒体系统。目前该计划只实现了 OggVorbis 这一部分。Ogg Vorbis 文件的扩展名是.OGG。这种文件的设计格式是非常先进的。现在创建的 OGG 文件可以在未来的任何播放器上播放,因此,这种文件格式可以不断地进行大小和音质的改良,而不影响旧有的编码器或播放器。

优点:

- Ogg Vorbis 的音质和 MP3 不相上下。由于 Vorbis 使用了与 MP3 相比完全不同的数学原理,因此在压缩音乐时受到的挑战也不同。在聆听测试中,同样位速率编码的 Vorbis 和 MP3 文件具有同等的声音质量。

- Ogg Vorbis 支持类似于 MP3 的 ID3 信息。Vorbis 格式中包括有一个灵活而又完整的注释栏，可用于填写各种相关信息。
- Ogg Vorbis 格式支持流式播放。音频流是 Vorbis 的一个重要组成部分。vorbis 格式从设计的一开始就是立足于可以容易地进行流式处理。并且，Vorbis 的设计者正与 Icecast 流媒体软件的创造者一起使 Icecast 兼容 Vorbis。
- Ogg Vorbis 避免了象 MP3 文件的 ID3 标记那样烦琐的操作。Vorbis 具有一个设计良好、灵活的注释，避免了象 MP3 文件的 ID3 标记那样烦琐的操作；Vorbis 还具有位速率缩放：可以不用重新编码便可调节文件的位速率。Vorbis 文件可以被分成小块并以样本粒度进行编辑；Vorbis 支持多通道；Vorbis 文件可以以逻辑方式相连接等。
- 缺点：没有提供帧格式、同步、及错误保护等。

(4) FLAC

格式说明：

FLAC 中文可解释为无损音频压缩编码。FLAC 是一套著名的自由音频压缩编码，其特点是无损压缩。不同于其他有损压缩编码如 MP3 及 AAC，它不会破坏任何原有的音频资讯，所以可以还原音乐光盘音质。2012 年以来它已被很多软件及硬件音频产品（如 CD 等）所支持。

优点：

FLAC 属于无损音频压缩文件格式（文件名为.flac）。FLAC 格式文件较小，便于播放。OGG 计划支持 FLAC，也可将 FLAC 音频数据封装在 OGG 传输层中，生成 OGG FLAC 格式文件（文件名为.ogg）。OGG FLAC 格式文件较大（比 OGG Vorbis 格式文件大得多），便于传输、编辑、播放。

该格式的特点是支持真实性验证技术、数码音质无损、文件格式开放透明、不绑定软硬件、文件格式自描述、格式自包、易于存储、信息安全。

缺点：

FLAC 仅仅是一个无损音频编码，不含有有损压缩的功能。因此当需要小体积有损编码的音频文件时，你将仍然需要 MP3、AAC、Ogg Vorbis 等编码；FLAC 不支持任何版权保护（防复制）方法，自行尝试对 FLAC 文件进行保护是徒劳无功的。

● 视频部分：

(1) AVI

格式说明：

AVI 英文全称为 Audio Video Interleaved，即音频视频交错格式，是微软公司于 1992 年 11 月推出、作为其 Windows 视频软件一部分的一种多媒体容器格式。AVI 文件将音频（语音）和视频（影像）数据包含在一个文件容器中，允许音视频同步回放。类似 DVD 视频格式，

AVI 文件支持多个音视频流。AVI 信息主要应用在多媒体光盘上，用来保存电视、电影等各种影像信息。

优点：AVI 是主流的视频格式，品质比 mov 等都清晰很多。图像质量好。

缺点：

AVI 文件比 mov, rmvb 偏大；AVI 有两种压缩格式 RMVB、WMV，压缩标准不统一。因此经常会遇到高版本 Windows 媒体播放器播放不了采用早期编码编辑的 AVI 格式视频，而低版本 Windows 媒体播放器又播放不了采用最新编码编辑的 AVI 格式视频。常用扩展名：.avi。

(2) MOV

格式说明：

MOV 即 QuickTime 影片格式，它是 Apple 公司开发的一种音频、视频文件格式，用于存储常用数字媒体类型。当选择 QuickTime (*.mov) 作为“保存类型”时，动画将保存为 .mov 文件。QuickTime 用于保存音频和视频信息，包括 Apple Mac OS,

Microsoft Windows 95/98/NT/2003/XP/VISTA, 甚至 WINDOWS 7 在内的所有主流电脑平台支持。

优点：

- MOV 是一种大家熟悉的流式视频格式，在某些方面它甚至比 WMV 和 RM 更优秀，并能被众多的多媒体编辑及视频处理软件所支持，用 MOV 格式来保存影片是一个非常好的选择。
- MOV 格式的视频文件小。
- MOV 格式的视频画面效果较 AVI 较好。

缺点：压缩之后内存变小但是牺牲了质量。

(3) MPEG

格式说明：

MPEG 标准主要有以下五个，MPEG-1、MPEG-2、MPEG-4、MPEG-7 及 MPEG-21 等。该专家组建于 1988 年，专门负责为 CD 建立视频和音频标准，而成员都是为视频、音频及系统领域的技术专家。及后，他们成功将声音和影像的记录脱离了传统的模拟方式，建立了 ISO/IEC 11172 压缩编码标准，并制定出 MPEG-格式，令视听传播方面进入了数码化时代。

优点：

特别针对低带宽等条件设计算法，因而 MPEG-4 的压缩比更高，使低码率的视频传输成为可能。在公用电话线上可以连续传输视频，并能保持图像的质量，这是其它技术做不到的。节省存储空间。同等条件如场景、图像格式和压缩分辨率条件下，经过编码处理的图像文件越小，所占用的存储空间越小。由于 MPEG-4 算法较 MPEG-1、MPEG-2 更为优化，因而在压缩效率上更高。

图像质量好。MPEG-4 的最高图像清晰度为 768X576，远优于 MPEG1 的 352X288，可以达

到接近 DVD 的画面效果。这使得它的图像高清晰度非常好。另外，其它的压缩技术由于算法上的局限，在画面中出现快速运动的人或物体和大幅度的场景变化时，图像质量下降。而 MPEG-4 采用基于对象的识别编码模式，从而保证良好的清晰度。

缺点：MPEG-1 频宽要求高。

2.5 图像文件保存格式探析

(1) TIFF

TIFF (Tag Image File Format) 即有标签的图像文件格式，是一种国际性的图像格式，它适用于大部分图像处理软件和图像浏览软件，在图形媒体之间的交换效率很高，且与硬件无关，是跨越 Mac 与 PC 平台最广泛的图像打印格式。一个独立的 TIFF 文件有很高的使用价值，它既可以作为页面上的元素，也可以用来直接输出到数字印刷机或激光照排机，因此，TIFF 通常用于打印和印刷输出的最终格式。目前很多数码相机都支持 TIFF 输出。

优点：

- 质量好而且兼容性高，不会受到处理软件的限制。
- 适用性强，格式受到广泛支持，尤其是在 Macintosh 计算机和基于 Windows 的计算机之间。
- 支持可选压缩。TIFF 最大的特点就是可以进行无损压缩，TIFF 既可以通过 LZW 进行内部压缩，又可以通过 WinZip 等软件进行外部压缩。
- 可扩展格式支持许多可选功能。

缺点：

- 图像文件大且在需要更多存储的时间。
- 该类格式的图像不受 Web 浏览器支持。
- 可扩展性会产生导致许多不同类型的 TIFF 图片，但并不是所有 TIFF 文件都与所有支持基本 TIFF 标准的程序兼容。

(2) JPEG2000

JPEG2000 正式名称为“ISO 15444”，是由 JPEG 组织负责制定的新的静止图像压缩标准。与 JPEG 不同，JPEG2000 利用当前最新的嵌入式编码技术，采用以小波转换(Wavelet transform)为主的多解析编码方式，生成的码流有较强的功能，可应用于多个领域，压缩效果优于 JPEG。

JPEG2000 目标是通过进一步改进目前压缩算法的性能以适应低宽带、高噪声的环境，以及医疗、图书或网上服务、保安等多方面的应用。它适用于各种类型的静止图像，如二值图像、灰度图像和多分量图(彩色图像)；支持自然图像、合成图像、医学图像和文本，允许使用不同的图像模型，如图像图书馆，带限信道的实时传输等。

优点:

- 为图像数据的处理方式及传输次序提供了极大的灵活性,可以方便地实现渐进式传输。它先传输图像的大体轮廓然后逐步传输其他数据,不断地提高图像质量。这样图像就由朦胧到清晰显示出来,充分利用和节约有限的带宽。
- JPEG2000 既支持有损压缩,也支持无损压缩。其在各种比特率下以及无损情况下的图像压缩性能均优于 JPEG。特别是在极低比特率和高比特率下 这种优势显得尤为突出。
- JPEG2000 支持对希望的特定区域进行特别的压缩处理。它既可以指定图像上任意区域的压缩质量,还可以指定哪个部份先进行解压处理。这在大大降低图像尺寸方面起到很大作用。
- 在与 JPEG 类似质量的前提下, JPEG2000 的压缩率比 JPEG 高 20% ~ 40%。若将 JPEG 图片全部换成 JPEG2000 编码方式,在同样的网络带宽下,对于图片下载的等待时间将大大缩短。

缺点:

- JPEG2000 的编码计算复杂性明显高于 JPEG,这不利于 JPEG2000 的软件实现。
- 对于文本图像及合成图像, JPEG2000 的无损图像压缩性能明显劣于 JPEG-LS。

(3) JPEG

JPEG (Joint Photographic Experts Group) 即联合图像专家组,是由一个在国际标准组织下从事静态图像压缩标准制定的委员会制定的第一套国际静态图像压缩标准: IS010918-1。

JPEG 的适用范围非常广泛,几乎所有的图像浏览和编辑软件都可以打开它,也是 Internet 上最常见的文件格式之一。应用于数码照相机时, JPEG 格式可以把文件大小压缩至 10%~20%,相对于未经压缩的原图像, JPEG 格式的画质下降和细节损失并不明显。

优点:

- JPEG 格式兼容性比较好,几乎所有的工作平台和软件都可以读取编辑,在许多不同的领域得到了广泛的应用。
- JPEG 文件中的数码较少,文件尺寸也较小这就意味着对设备的要求比低。
- 能够满足大量高质量图像的存储量的要求。它采用的 DCT 变换编码和熵编码,具有适中的计算复杂性,易于硬件实现,由于它在保证图像质量的前提下能提供较高的压缩比,通过使用专用压缩芯片, JPEG 甚至可以用于较高波特率下连续图像的传输。
- 利用可变的压缩比可以控制文件大小。 JPEG 方法对于大多数类型的图像均能取得比较满意的编码压缩结果,由此可以节省大量的数据。

缺点:

- JPEG 不适用于所含颜色很少、具有大块颜色相近的区域或亮度差异十分明显的较简单

的图片。

- 在低比特率压缩时, JPEG 图像压缩技术中基于分块的 DCT 变换编码会产生方块效应, 严重影响解码图像的视觉效果。其主要原因是低比特率压缩的粗量化过程在各个方块内引起高频量化误差, 各子块独立编码而没有考虑块间的相关性, 从而造成块边缘的不连续性。
- 有损耗压缩导致原始图片数据质量下降。由于舍去了图像的高频信息, 因而编码图像的边缘难以很好地保持, 例如图像中有文字的地方, 这样的失真就显得很明显。
- 编辑并重新保存 JPEG 文件时, 其会混合原始图片数据的质量下降, 且这种下降是累积性的。

(4) GIF

GIF 格式(Graphics Interchange Format)即图形交换格式, 是由美国一家著名的在线信息服务机构 CompuServe 于上世纪 80 年代针对当时网络传输带宽的限制而开发的作为一个跨平台图形标准而开发的、与硬件无关的 8 位彩色文件格式, 也是目前在因特网上使用最早、应用最广泛的图像格式。GIF 格式有 87A 和 89A 两种版本, 其支持的颜色数目最多只有 256 种, 文件的大小取决于实际使用的颜色数, 采用 LZW 压缩编码。

GIF 格式还支持透明图像和动画。GIF 动画格式(89A 版本)可以同时存储若干幅静止图像并制定每幅图像轮流播放的时间, 从而形成连续的动画效果, 目前因特网上大量采用的彩色动画文件多为这种格式的 GIF 文件, 且很多图像浏览器都支持直接观看此类动画文件。

优点:

- GIF 格式压缩比高、磁盘空间占用较少, 且广泛支持 Internet 标准。
- 支持无损耗压缩和透明度。

使用广泛, 目前 Internet 上大量采用的彩色动画文件多为这种格式的文件。

- GIF 图像格式还增加了渐显方式, 为迎合用户实用心理, 考虑到网络传输中的实际情况, 在图像传输过程中, 用户可以先看到图像的大致轮廓, 然后随着传输过程的继续而逐步看清图像中的细节部分。
- GIF 图像文件短小、下载速度快, 且可由许多具有同样大小的图像文件组成动画文件。

缺点:

- GIF 格式只支持 256 色调色板, 因此, 详细的图片和写实摄影图像会丢失颜色信息。
- 在大多数情况下, GIF 格式的无损耗压缩效果不如 JPEG 格式或 PNG 格式。
- GIF 格式仅支持有限的透明度, 没有半透明效果或褪色效果)。

(5) PNG

PNG (Portable Network Graphic Format) 即“可移植网络图像文件格式”, 是一种与平台无关的新兴网络图像格式。PNG 格式支持 8 位调色板、16 位灰度级。而且可达到 48 位

RGB 的支持。GIF 格式图片以审核颜色深度储存单个光栅图像，采用无损压缩技术，使图像文件的质量不会因为压缩而缺损。

1994 年底，由于 Unysis 公司宣布 GIF 拥有专利的压缩方法，要求开发 GIF 软件的作者须缴交一定费用，由此促使免费的 png 图像格式的诞生。PNG 一开始便结合 GIF 及 JPG 两家之长，打算一举取代这两种格式。1996 年 10 月 1 日由 PNG 向国际网络联盟提出并得到推荐认可标准，并且大部分绘图软件和浏览器开始支持 PNG 图像浏览，从此 PNG 图像格式生机焕发。

优点：

- PNG 格式兼有 GIF 和 JPEG 的色彩模式，可以定义 256 个透明层次，并通过这 256 个透明层次可以实现多种特殊的图像效果。
- PNG 格式支持高级别无损压缩，其采用基于 LZ77 的无损压缩，8 位的 PNG 图像比起相同的 GIF 图像小 10%-30%，同时它还可以在图像文件中存储一些文本注释信息。
- PNG 格式支持 alpha 通道透明度和伽玛校正。
- 大多数情况下，最新的 Web 浏览器支持该格式。

缺点：

- 对硬件要求较高，较旧的浏览器和程序可能不支持 PNG 文件。
- 作为 Internet 文件格式，与 JPEG 的有损耗压缩相比，PNG 提供的压缩量较少。
- 支持文件类型有一定局限性，PNG 格式对多图像文件或动画文件不提供任何支持。

(6) DjVu

DjVu 是由美国 AT&T 实验室于 1996 年开发研究的一种使得高质量的扫描图像在因特网上的发布更为容易的图像压缩技术。其主要技术是将图像分为背景层(纸的纹理和图片)、前景层(文本和线条)和掩码层(形状库)，通过分离文字和背景，DjVu 可以用高分辨率来还原文字，使锐利边缘得以保留，并最大限度地提高可辨性；同时用较低的分辨率来压缩背景图片，从而使整个图像的质量得到保证。

为了适应网络时代图像资源的发展需求，DjVu 针对图像的不同图层，采用了 JB2 与 IW44 等不同的先进图像压缩技术，独特的双层格式使其兼具图像与文档的特性——既能满足图像浏览的要求，又可实现全文检索的目标。其先进的设计理念与处理技术，使浏览与下载变得非常便捷。不同于以往其它图像格式，DjVu 无需安装庞大的图像处理软件，只需下载安装适用于不同浏览器的免费插件 DjVuBrowserPlug-in 即可。

优点：

- 具有清晰度高、压缩比高、显示效果好、传输速度快、生成文件小等特点，便于保存利用。
- 支持双层格式与文本检索。DjVu 图像不同与其他格式文件，在生成 DjVu 图像时，可同

时进行 OCR 识别，直接生成双层格式 DjVu 文件，也可通过专业软件，将其他格式文件或 DjVu 单层文件进行 OCR 识别，转换成双层格式 DjVu 文件。更具不同的是图像与文本在同一个文件中，既保持原始图像的面貌又可实现全文检索与复制。

- 操作方便。DjVu 浏览器具备放大镜功能，用户在浏览 DjVu 图像时只需按 Ctrl 键即可实现局部放大，省去将整幅图像放大与拖拽浏览的麻烦。
- 输出方式灵活。在保存 DjVu 文件时，也可根据需要灵活进行：既可整页保存，也可选择区域进行局部复制，对于多页文件还可选择以多页文件形式或多个单页文件形式输出。
- 安全性较高。DjVu 格式能够与数字水印系统相结合，保证数字文件的真实性及其在网络中的合法使用。其可以将一个标识图像与宿主图像数据进行叠加，形成带可见水印的图像，或者将一个特定的序列号经过加密算法嵌入宿主图像数据，得到不可见水印。
- 适用范围广泛。DjVu 被广泛应用于目录、图书、杂志、照片、手稿、地图、艺术画、古籍、书法等各类图像数字资源。

3. 流式文档与版式文档

电子文档根据不同的存储方式一般可分为流式文档和版式文档。流式文档主要存储的是逻辑数据，文字内容主要借助流式布局进行从上到下自然、排版，受不同环境影响，呈现不同排版效果。流式文档的特点是可排版，能适应当前的显示环境，在编辑、用户交流文档等方面优势突出。流式文档格式以微软的文字处理格式 DOC 为代表。版式文档是将文字、图形、图像等多种数字内容对象按照一定规则进行版面固化呈现，是一种独立于软件、硬件、操作系统、呈现/打印设备的电子文档格式。版式文档主要应用于成文后文件的发布、传播和存档。版式文档的特点是版本固定、不跑版所见即所得。典型版式文档格式有 PDF，以及我国自主制定的标准版式文档格式 OFD。

流式文档、版式文档各有自己的优势。流式文档易于编辑，版式文档则能与原文档版面保持一致。

4. 国外长久保存战略的启示

4.1 澳大利亚电子文件长期保存策略对我国的启示

(1) 制定电子文件长期保存国家战略

电子文件长期保存需要有一个全盘考虑、系统规划的顶层设计。我国需要建立电子文件长期保存战略,明确国家电子文件长期保存工作的基本态度和总体思路,并通过一系列基本制度的安排与实施,实现电子文件的长期保存,增强信息资源的国家掌控能力。

我国应该将电子文件长期保存纳入国家信息化发展规划,纳入各行各业业务系统建设,在国家层面以及行业层面上给予支持,引导电子文件保存事业有步骤地开展,防止各自为政、各搞一套的混乱局面出现。

(2) 完善管理机制

完善国家电子文件长期保存管理机制,有利于在全国范围内有组织有计划地构建一个国家电子文件长期保存协作网,最全面、最经济、最有效地开展工作。

完善国家电子文件长期保存管理机制,需要以国家档案局为主,同时联合各部委、各省市的信息管理部门、技术支持部门、智力机构等诸多部门,构建国家电子文件长期保存体系,明确相关管理机构的责任,强化管理协调机制和审计监督机制,从法规、经济、方针政策和行政管理等方面来支撑和约束。

同时,我国可以借鉴澳大利亚的实践模式,建立国际间、国家级别机构之间、行业内、地区间的合作机制,整合更多的社会力量,应对电子文件长期保存的诸多挑战。

(3) 推动法律法规建设

从澳大利亚情况来看,立法是推动电子文件长期保存工作最基本的保障。我国必须加快相关立法工作,使电子文件长期保存有法可依、有章可循。

澳大利亚除了颁布有利于长期保存的档案法或著作权法案之外,还基于文件是各项业务活动的工具与证据、文件管理是各项业务活动有机组成部分的认识,在刑法、行政法、民商法等其他众多法律中列有许多条款涉及文件管理的要求。而我国很少在其他法律中提及文件管理问题。当前情况下,修订其他法律的可行性很小,因此应该重点制定一部较为完善的《中华人民共和国文件法》。

《中华人民共和国文件法》应该包括以下内容:(1)明确认定电子文件属于文件的范畴,规定满足一定条件的电子文件具有证据效力;(2)明确赋予档案部门全过程监控文件管理的职责;(3)明确档案部门、业务部门、信息部门及其他有关部门相应的文件管理义务与权利;(4)规定文件生成、传递、归档、保存和开发利用的基本宗旨和原则;(5)规定国家文件管理的体制与运行机制。

(4) 制定可实施的专门政策

电子文件长期保存是一项利国利民的公益行动,需要国家的支持,需要全社会的共同参

与。我国亟需制定电子文件长期保存专门政策，以表明国家在这方面的意志和价值观，体现国家的支持；以有效引导社会正视、重视长期保存工作，提高公众和机构的数字保存意识。

同时，要注意保持专门政策与国家数字信息管理和档案管理政策的一致性。电子文件长期保存政策应该是政府数字信息管理整体措施的一部分，与其他政策措施共同推进政府数字信息管理工作。如，澳大利亚国家馆在制定《数字连续性计划》之后，发布了《政府数字转型政策》，两者在流程上前后衔接，内容上互为支撑。

（5）构建有利于电子文件长期保存的标准、规范体系

澳大利亚坚持文件连续体理论，制定了覆盖文件全生命周期的标准和规范体系，这些标准和规范相互补充、相互融合，为澳大利亚实现电子文件真实、完整与长期可读提供了强有力的保障。

我国在构建电子文件长期保存的标准、规范体系过程中，要充分利用国内外先进理论，借鉴已有国际标准，充分考虑电子文件全生命周期管理需要，建立电子文件保存标准、规范体系的制定目标，联合社会多种力量共同参与，全面提高相关领域标准化的整体水平，加强与电子文件长期保存相关的标准体系与规范体系建设，加强有利于标准实施的标准化策略手段建设和相关机制建设，选取有利于系统开放性和扩展性的标准加以推荐。

考虑到电子文件国家标准体系的制定过程会持续很长时间，目前可以先制定一个较为全面、系统、先进的《电子文件长期保存管理指南》，暂时作为全国电子文件长期保存的政策引导和规范。

（6）重视技术研发

载体损伤和技术老化是电子文件长期保存面对的最直接的挑战，因此技术问题在澳大利亚备受重视。澳大利亚国家馆通过开放源代码方法，于2004年开发了Xena 1.0版，2009年升级至Xena 4.2.1版，并建立了数字保存软件平台。

我国要明确技术研究路线与研究方向，增加技术研究动力，形成有利于相关信息技术研究和普及的良性机制，重点是在相关核心技术的研发中取得突破，在各行各业重点部门与核心部门的整体技术解决方案的研究开发方面取得商家成效，发挥市场机制、监督审计机制、最佳实践示范推广机制、相关技术市场培育机制和监督管理机制在技术推广普及中的作用。

我国在技术研发推广过程中，要充分考虑电子文件长期保存技术的易用性和兼容性；使用开放源代码技术，提高透明度、保障真实性、保证同行的参与和合作；同时要不断升级，保持技术的先进性。

（7）积极开展实践项目

澳大利亚电子文件长期保存启动项目众多，涉及的研究领域广泛，从保存战略规划、保存技术的开发和应用到保存元数据和规范的制定都做了大量的研究和实践。

我国需要从电子文件长期保存的全局角度出发，借鉴澳大利亚相关经验，整体规划实践

项目。国家可以选择前期基础较好、研究能力较强的单位，给予必要的资金和政策支持，有针对性的研究长期保存中的管理、技术、标准、人才等基本问题与关键问题，探寻最佳解决方案，推行最佳实践，实现示范效应。

4.2 国外标准化组织、美国国家航空航天局、美国归档特别工作组、美国国家档案与文件署、英国公共文件署、澳大利亚国家档案馆的几点启示

近年来，国内不少学者翻译介绍了大量国外重大电子文件长期保存项目的研究成果，在这个过程中也引进了很多新的概念和思想，这使我国电子文件长期保存研究积累了较为丰富的学术资源，但这还不足以解决目前我国电子文件长期保存的相关问题，我国的电子文件长期保存研究要做到在立足本国实际的同时，合理应用国外先进成果和经验。

(1) 制定电子文件长期保存国家战略

从目前来看，与国外先进国家项目相比，我国电子文件保存与管理工作具有较强的自发性和分散性，缺少国家层面的整体设计、统一规划、战略部署与组织领导，很难保证整个社会电子文件的科学管理、完整保存和有效利用。制定电子文件长期保存国家战略，旨在建立全国范围电子文件安全管理、长期保存和有效利用的网络体系，这不仅对电子文件长期保存项目的开展至关重要，对于电子文件管理实践的顺利推进也同样举足轻重。

(1) 建立电子文件长期保存组织保障机制

国外先进国家的发展经验告诉我们，电子文件长期保存研究必须取得政府部门及相关机构的有力支持，因为电子文件长期保存体系的建立需要有人才、资金、技术、管理体制、政策规范等支撑，而这些都依赖于科学系统的组织保障措施，以保证自上而下地建立电子文件长期保存体系，并确保该体系的正常运转。至少应从两个方面考虑组织机制的构建：一方面，确立一个权威的国家层面的领导机构负责有关电子文件管理国家战略、方针政策的制定，领导和协调全国电子文件长期保存工作；另一方面成立各行业、各领域、各层级信息机构的综合协调部门，加强各领域间合作，并给予一定政策、经济、技术和智力上的支持，便于相互借鉴，共同发展。这样，才能保证我国电子文件长期保存理论与实践健康有序地发展，构建电子文件长期保存网络体系。

(3) 建立电子文件长期保存长效合作机制

电子文件长期保存工作具有广泛的社会性，不仅涉及各个行业、各类人员，还涉及技术、标准、法律等众多研究方向，需要合理有效的协作，因此需要在具备组织保障的前提下，建立广泛长效的合作研究机制。国外很多项目都基于合作，包括跨部门、跨行业，甚至跨国合作，而我国电子文件长期保存研究是局部的、分散的，虽然也有国内及国际合作项目，但国内研究规模都很有限。因此，应从以下几方面着手，建立合作机制。第一，应加强地区合作，有利于在电子文件长期保存项目研究中相互借鉴，取长补短，更有利于发现关键问题，集中优势力量攻克难题。第二，与技术部门合作，有利于采取先进的技术手段策略，开展研究。

第三，与相关文献信息机构合作，在我国，图书馆界对于数字资源长期保存研究要相对深入和广泛，在进行电子文件长期保存问题研究时，可以充分借鉴吸收他们的成功经验，少走弯路。第四，加强国际交流与合作，不断完善我国电子文件保存策略和技术。通过不同方式、不同领域的合作，有利于实现技术的合作攻关、资源的合理分工、经费的最佳分配。

(4) 构建电子文件长期保存标准规范体系

标准规范体系的构建是推动电子文件长期保存理论与实践发展的重要动力之一。从长远考虑，电子文件的长期保存和有效利用必须考虑标准化和规范化问题。我国电子文件长期保存的标准制定工作相对滞后，虽然早在十多年前就先后发布了《CAD 电子文件光盘存储、归档与档案管理要求》(GB/T17678.1/2-1999)与《电子文件归档与管理规范》

(GB/T18894-2002)两个国家标准，但后续研究一直不成体系，也没有形成广泛接受和遵循的国家或行业标准规范。因此，现阶段应根据我国电子文件管理工作实际，以 OAIS 等相关国际标准为基本遵循，借鉴国外电子文件长期保存先进经验和技术，抓紧制定我国电子文件长期保存标准，构建标准规范体系。我国电子文件长期保存标准规范体系至少应包括：系统需求相关标准、元数据相关标准、数据传输相关标准、数据存储相关标准、日常维护相关标准、信息安全相关标准和电子文件保存操作规范等。

(5) 构建电子文件长期保存技术保障体系

电子文件长期保存是一项十分艰巨而复杂的工程，涉及方方面面，会遇到很多困难和技术问题，如系统结构、存储技术、档案系统功能需求等等，都需要最先进的技术作为基本支撑。从国外一些成功案例可以看出，电子文件长期保存项目的顺利开展依赖于科学的技术方案和策略，构建电子文件长期保存技术保障体系，从功能需求、保存能力、保存内容、保存技术、保存方式等几个方面设计保存方案和策略，并综合考虑保存系统对各种类型文件的包容性，对各类操作系统及软件的兼容性，对各类操作人员的实用性，以及对用户的易用性。其中，功能需求设计是最为关键和基础的工作，很多成功项目的功能需求分析都十分详尽系统；保存能力分析主要依据项目规模，整体规划完成项目所必须的人力、物力等软硬件条件；保存内容规划是对保存对象的全面分析，包括格式、数量、价值等等；保存技术和保存方式的选取是根据保存对象的特点，采取最合适的保存技术和方式，如：封装、复制、仿真、迁移以及光盘、磁盘阵列、网络存储器等。

4.3 美国电子文件长期保存策略对我国的启示

各种不同类型电子文件保存的需求应该确定下来，同时应该尊重档案的原始性以及档案卷宗的一致性。Duranti 教授主持的 InterPARES 项目就集中了来自全世界的高校、档案机构的档案专家以及计算机、信息技术科学家和工程师一起协商明确保存真实的电子档案的需求。InterPARES 项目从真实性上定义了电子文件保存的需求。同时，InterPARES 项目还检查了与电子文件保存相关的技术问题并建立了一个正式的电子文件保存模型。这些工

作在进行的同时，一些想法被提出来讨论。一个很关键的问题是，从严格意义上来说，保存真实的电子文件是不可能的；只能说是保存了一种再造电子文件的能力。将以二进制代码形式保存在存储设备上的电子文件恢复成并通过软件得以呈现或传递，这种能力在电子文件保存过程中必须得到保证。类似的，音乐唱片保存的不是音乐本身，只不过保存了有一些象征性的音乐号，当使用恰当的装置播放时，就能播放出音乐。如果保存电子文件的程序执行正确的话，程序的输出结果就应该是记录，而不是电子文件的保存形式——二进制文件。这是一个很重要的理念。这个理念将使你在保存电子文件时考虑问题的重点放在存储、输出、传递电子文件这整个过程上，而不是紧盯着存储本身不放。电子文件必须被不可避免的被重现的理念对强调论证电子文件的一致性和真实性非常重要。这个理念将传统的完整的监管链的概念延伸到了电子文件的完整保存过程的概念上了。ICA 手册中明确定义：电子文件必须在满足以下条件才能被保存：电子文件能长久以它被恢复的形式存在，并且一旦恢复，必须恢复成可靠的真实的形式。

与 InterPARES 项目的进行的同时，NARA 也主持了一项信息管理架构的研究，这个信息管理架构也是致力于电子文件保存需求的，也就是 ERA 的架构，ERA 的架构采用了符合 ISO 的标准的 OAIS 架构。OAIS 机构在电子文件保存需求方面扩展了一般通用的参考模型。为了应对由于技术更新带来的基本问题，这种架构假定 ERA 是独立于任何特定信息技术。也就是说，ERA 系统的任何一个软件或硬件模块都可以随意替换，而不影响整个系统的运作。随着信息技术高速发展，电子文件保存问题日益严峻，美国国家档案馆电子文件档案馆项目从启动、建设到使用经历了十几年，积累了一系列的经验，取得丰硕的成果。相比而言，我国电子文件的研究有较大差距，通过 EAR 电子文件相关经验的学习研究，对我国电子文件的研究进展有一定的帮助和借鉴作用。

5.总结

5.1 选择长久保存格式时要考虑的因素：

由于电子档案具有不同的数据类型，因此，很难确定一个统一认可的格式选择标准，选择什么样的格式保存电子档案是由电子档案创建过程中的操作要求和电子档案长期存取的要求所决定的。综合各国选择标准以及国内有关规定和各位学者的观点后，笔者认为：对电子档案长期保存格式的选择标准的必然要求是安全可靠、具有高保真性并且是不可更改的。我们可以参考以下几点：

(1) 开放性。即格式开放，有公开发表能够免费获取的相应标准和技术规范，有与产品无关的技术专家组和标准化组织支持该格式。这种开放性的格式减少了因为某些公司倒闭而影响电子档案读取利用的风险，并且减少了跨平台迁移的风险和费用。

(2) 标准性。即该格式被广泛使用，具有自由使用的技术规范，可转换性高，支持保存元数据。这样在长期保存电子档案时，就能够做到降低仿真、迁移的频率。

(3) 稳定性。不会随着时间的推移而作重大的变更，而且新版本的格式应该能够做到兼容不同的版本（向上兼容和向下兼容）。

(4) 广采性。数字格式的广采性包括个人电脑随机软件包、Web 浏览器的本地化支持、数字产品格式易转换等。软件厂商对创建良好和广泛使用的格式往往要比仅仅瞄准新市场的格式具有更宽泛和更长久的支持。流行格式会尽可能多地受到软件厂商的支持，人们可能更喜欢采用。

(5) 软硬件平台的独立性。被多种操作系统和应用软件支持。支持多种存储技术，或与存储技术无关。当用户不能使用指定产品软件时，可使用已有的插件读取。使用与设备无关的颜色规范实现准确打印和再现，不必考虑软硬件平台。

(6) 真实性。电子文件的真实性要求保留“原貌”（例如字体、颜色和外观），数据迁移要求选取的文件格式能够保留文件“原貌”。

(7) 可利用性。要确保电子文件可读、可利用。综上所述，在电子文档长期保存格式的选择时，我们要从保存需求分析入手，根据保存格式的选择标准，综合分析各种格式的利弊，最终选择最适于长期保存的文档格式，以保证电子文档的真实性和长期可利用性。

5.2 长期存储格式的相关规定

目前，我国对于电子档案长期保存格式选择的相关规定主要有以下几个：

《电子文件归档与管理规范》

《电子文件归档与管理规范》中对各种类型的电子文件可采用的文件格式均作出了规定。具体如下：

“对用文字处理技术形成的文本电子文件，收集时应注明文件存储格式、文字处理工具等，必要时同时保留文字处理工具软件。文字型电子文件以 XML、RTF、TXT 为通用格式。”

“对用扫描仪等设备获得的采用非通用文件格式的图像电子文件，收集时应将其转换成通用格式，如无法转换，则应将相关软件一并收集。扫描型电子文件以 JPEG、TIFF 为通用格式。”

“对用计算机辅助设计或绘图等设备获得的图形电子文件，收集时应注明其软硬件环境和相关数据。”

“对用视频或多媒体设备获得的文件以及用超媒体链接技术制作的文件，应同时收集其非通用格式的压缩算法和相关软件。视频和多媒体电子文件以 MPEG、AVI 为通用格式。”

“对用音频设备获得的声音文件，应同时收集其属性标识、参数和非通用格式的相关软件。音频电子文件以 WAV、MP3 为通用格式。”

《纸质档案数字化技术规范》

国家档案局于 2005 年 4 月 30 日发布了《纸质档案数字化技术规范》，其中规定对纸质档案进行扫描时，“采用黑白二值模式扫描的图像文件，一般采用 TIFF(G4)格式存储；采用灰度模式和彩色模式扫描的文件，一般采用 JPEG 格式存储。存储时的压缩率的选择，应以保

证扫描的图像清晰可读的前提下，尽量减小存储容量为准则”。“提供网络查询的扫描图像，也可存储为 CEB、PDF 或其他格式”。“目录建库应选择通用的数据格式，所选定的数据格式应能直接或间接通过 XML 文档进行数据交换”。

《版式电子文件长期保存格式需求》

2009 年，安徽省档案局牵头制订了开放的版式文件存档格式标准《版式电子文件长期保存格式需求》(DA/T47—2009)，系统地界定了版式电子文件长期保存格式必须满足的 11 个方面的需求。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。

5.3 适合长期保存的格式

(1) PDF/A 格式

PDF 格式规范已经公开,并可免费获取,具有广泛的支持性。该格式具有与平台、字体无关的特点,解除了电子文件与操作系统的依赖关系,用户可以通过它的通用浏览器 Adobe Reader 在 Windows、UNIX、Mac 等任一系统平台上原版原式地读取 PDF 文档,不受操作系统、网络环境、应用程序的版本、字体的限制,从而从电子文件自身的角度保证了电子文件的可读性。PDF 格式所具有的高度兼容性使用户可以轻松地将任何文档原版原式地转换成 PDF 文档,进而固化信息,使之不易更改,保持了原文档信息内容和结构的完整、真实,为不同软件生成的电子文件提供了一个可完整迁移其信息内容的统一文件格式。

在各国政府机构包括美国、英国、德国、新加坡、印度、澳大利亚等的电子政务领域中,PDF 格式也被广泛应用。综合来说,PDF/A 格式的特点是支持数字签名、格式开放、不绑定软硬件、格式自包含、格式自描述、固定显示、不包含加密、可向其他文本格式转换等。

(2) TIFF 格式

TIFF(标记图像文件格式)是由 Aldus 和 Microsoft 公司为桌面出版系统研制开发的一种灵活的位图图像格式,用于应用程序之间和计算机平台之间交换文件。实际上已被所有绘画、图像编辑和页面排版应用程序所支持,而且几乎所有桌面扫描仪都可以生成 TIFF 图像。

TIFF 支持多种编码方法,其中包括 RGB 无压缩、RLE 压缩、LZW 压缩、ZIP 压缩、CCITT 压缩等,6.0 版本又增加了 JPEG 压缩选项。TIFF 格式由于存储图像质量高非常有利于原稿的复制。另外,所有绘画、图像编辑和页面排版应用程序以及大多数扫描仪对 TIFF 格式都提供良好的支持,这使得 TIFF 格式成为数字图像处理的最好选择。

(3) JPEG2000 格式

JPEG2000 于 2001 年正式成为国际标准。JPEG2000 完全采用小波变换代替余弦变换,与 JPEG 相比,具备了更高的压缩比以及更多新功能。在文件大小相同的情况下,JPEG2000 压缩的图像比 JPEG 质量更高,其压缩率比 JPEG 高 30%左右,精度损失更小。

(4) FLAC 格式

FLAC 属于无损音频压缩文件格式(文件名为.flac)。FLAC 格式文件较小,便于播放。

OGG 计划支持 FLAC,也可将 FLAC 音频数据封装在 OGG 传输层中,生成 OGG FLAC 格式文件(文件名为.ogg)。OGG FLAC 格式文件较大(比 OGG Vorbis 格式文件大得多),便于传输、编辑、播放。该格式的特点是支持真实性验证技术、数码音质无损、文件格式开放透明、不绑定软硬件、文件格式自描述、格式自包、易于存储、信息安全。

(5) MPEG 格式

MPEG 即 Moving Picture Expert Group(运动图像专家组),用该编码器编码的影音文件扩展名可以是 MPG、MPE、MP4、M1V、M2V 等,其中 M1V 和 M2V 都表示该影音文件中不包含音频部分,只有视频部分。这种格式也流行了很长时间。未压缩的 AVI 格式文件经 MPEG-1、MPEG-2 和 MEG-4 等编码压缩后,图像容量大幅度缩小,画质则降低不大,而且图像质量视编码参数而定。该格式具有压缩算法公开、不绑定软硬件、易于转换等特点。

参考文献:

- [1] 王树亮,王巧英.图像压缩与 JPEG2000[J].江苏理工学院学报,2014,(4):8-11.DOI:10.3969/j.issn.1674-8522.2014.04.002.
- [2] 吴猛.基于 Web 的数据可视化技术初探[J].福建电脑,2007,(12):58-59.DOI:10.3969/j.issn.1673-2782.2007.12.033.
- [3] 李章晶,郑国勤.针对无线通信领域的图像压缩的研究[J].计算机工程与设计,2006,27(23):4471-4474.DOI:10.3969/j.issn.1000-7024.2006.23.028.
- [4] 范家如.浅谈电子文件长期保存格式 DjVu[J].卫生职业教育,2009,27(23):41-42.DOI:10.3969/j.issn.1671-1246.2009.23.021.
- [5] 刘聪明,郑晓惠.DjVu 在资源建设中的应用及其探讨[J].现代情报,2008,28(7):85-87.DOI:10.3969/j.issn.1008-0821.2008.07.029.
- [6] 梁民,王北.DJVU 格式与 PDF 格式的比较[J].电脑知识与技术,2011,07(16):3858-3859.DOI:10.3969/j.issn.1009-3044.2011.16.042.
- [7] 颜丽.GIF 和 JPEG 格式哪个好[J].科学 24 小时,2003,(10):44-45.
- [8] 丁绪星,朱日宏,李建欣等.JPEG2000 及其新技术[J].计算机应用,2003,23(2):1-4.
- [9] 赵锦,邢长征.JPEG2000 的编码原理及其优缺点[J].辽宁工程技术大学学报(自然科学版),2004,23(z1):192-193.DOI:10.3969/j.issn.1008-0562.2004.z1.075.
- [10] 杨小帆,何中市,柏森等.JPEG2000 的编码原理及其优缺点[J].计算机科学,2002,29(4):82-83.DOI:10.3969/j.issn.1002-137X.2002.04.026.
- [11] 许书道.PDF 和 DjVu 电子文档格式的分析与比较[J].数字技术与应用,2010,(1):121-122.
- [12] 南洋,王家民.RAW 格式与 JPEG 格式比较分析研究[J].电影评介,2012,(10):81-82.DOI:10.3969/j.issn.1002-6916.2012.10.037.
- [13] 王毅栋.数码照片常用格式[J].科技情报开发与经济,2007,17(5):233-234.DOI:10.3969/j.issn.1005-6033.2007.05.135.
- [14] 李秀.计算机文化基础(第5版)[M].北京:清华大学出版社,2005.