

## 大学数字图书馆国际合作计划

### 音频资料数据加工标准与操作规范

分类： 数字对象加工标准规范集

编号： jg\_zz\_audio

负责人： 孙红杰（上海师范大学）

完成日期： 2011年09月

## 文档备忘录

编号	标题	日期	状态
1	音频资料数据加工标准与操作规范	2011.09	成果草案

(标准规范在正式发布前后, 根据意见会有几次主要修订, 在此留下记录信息)。

## 版权声明

本报告研究工作属于教育部项目“大学数字图书馆国际合作计划”的一部分。按照有关规定，国家和“大学数字图书馆国际合作计划”项目管理中心拥有本报告的版权，依照《中华人民共和国著作权法》享有著作权。

本报告可以复制、转载、或在电子信息系统上做镜像，但在复制、转载或镜像时须注明真实作者和完整出处，并在明显地方标明“大学数字图书馆国际合作计划项目资助”的字样。

报告版权人不承担用户在使用本作品内容时可能造成的任何实际或预计的损失。

本报告作者谨保证本作品中出现的文字、图片、声音、剪辑和文后参考文献等内容的真实性与可靠性，愿按照《中华人民共和国著作权法》，承担本作品发布过程中的责任和义务。教育部、“大学数字图书馆国际合作计划”项目管理中心等有关管理机构对于本作品内容所引发的版权、署名权的异议、纠纷不承担任何责任。

“大学数字图书馆国际合作计划”项目组网站 (<http://www.cadal.cn>) 是本报告的第一发表单位。在不发生重复授权的前提下，报告撰写人保留将经过修改的项目成果向正式学术媒体直接投稿的权利。

## 摘要

本课题隶属教育部 211 重点工程项目“大学数字图书馆国际合作计划”（CADAL）的二期建设内容，它旨在为音频资料的数据加工和操作过程提供一个参考性方案，其中对关于本课题的若干术语进行了定义和解释，对音频数据加工的操作流程做了详细的交代，区分了数字资源的两种应用级别（备案保存与发布共享），并分别制定了这两种级别各自的数据加工标准，此外还对数字化之后音频文件的命名方式进行了建议。

# 内容目录

<b>1 概述</b>	
1.1 建设的背景.....	1
1.2 课题的基础.....	1
1.3 建设的内容与目标.....	1
1.4 本规范的应用范围.....	1
<b>2 术语定义</b> .....	1
2.1 数字音频.....	1
2.2 采用率.....	2
2.3 量化级.....	2
2.4 声道数.....	2
2.5 信噪比.....	2
2.6 比特率.....	3
2.7 杜比降噪.....	3
2.8 无损压缩.....	3
2.9 有损压缩.....	3
2.10 备案保存.....	3
2.11 发布共享.....	3
<b>3 标准规范内容</b> .....	4
3.1 制定规范的基本原则.....	4
3.2 数字加工的操作流程.....	4
3.2.1 资源整理.....	4
3.2.2 信息采集.....	4
3.2.3 格式选择.....	5
3.2.3.1 备案保存级.....	5
3.2.3.2 发布共享级.....	6
3.2.4 音频编辑.....	7
3.2.5 附件处理.....	7
3.2.6 元数据处理.....	7

3.2.6.1 元数据著录.....	7
3.2.6.2 元数据关联.....	7
3.2.7 文件命名.....	7
3.2.8 成果保存.....	8
3.2.9 质量检测.....	8
<b>4 参考文献</b> .....	<b>9</b>
<b>5 附录</b>	
5.1 资源类别代码表.....	10

# 1 概述

## 1.1 建设的背景

本课题隶属教育部 211 重点工程项目“大学数字图书馆国际合作计划”（CADAL）的二期建设内容，本期建设对一期数字化资源（以图书资源为主）的对象类型做了拓展，进而纳入了音、视频文件等媒体资源，这在提升 CADAL 资源平台的品格、优化其知识结构、拓宽其受众层面等方面具有显而易见的积极意义。

## 1.2 课题的基础

本规范在制定过程中参考了我国数字图书馆标准规范建设课题成果《通用数字资源（音频数据）格式标准分析报告》、《数字资源加工标准与操作指南》，国家图书馆对象数据项目成果《音频数据加工标准与工作规范标准规范》，大学数字图书馆国际合作项目成果《音频资料描述元数据标准规范》（附著录细则建议），以及科技部基础性工作专项资金项目成果《音频资料资源分析报告》、《音频资料描述元数据规范》等现有成果。

## 1.3 建设的内容与目标

本规范旨在为音频资料的数据加工和操作过程提供一个参考性方案，其建设内容包括：界定相关术语的含义，交代数据加工的操作流程，制定数据加工的标准，规定数字化之后音频文件的命名方式，等。

## 1.4 本规范的适用范围

本规范适用于盒式录磁带、开盘式录音磁带、数字录音带（DAT）、胶质粗纹唱片（S.P.）、胶质密纹唱片（L.P.）、激光唱盘（CD）、MP3 等模拟性音频原生资源、数字性音频原生资源及数字音频格式资源的信息采集、编码转换与保存发布。

# 2 术语定义

## 2.1 数字音频

数字音频（Digital Audio）是一种利用数字化手段对声音进行录制、存放、编辑、压缩或播放的技术，它根据一定的采样率、比特率、压缩率和量化级等参数，对模拟信号进行采样，最终转化为以二进制数据保存的数字文件。这些数据信号可被再次转换为模拟的电平信号来实现播放。就存储方式而言，数字声音与一般的磁带、广播、电视里的声音有本质区别，后者是其主要的信息来源。数字音频的主要传播媒介是激光唱盘和网络，在网络通讯、远程培训、数字图书馆等领域有重要的应用价值。

## 2.2 采样率

所谓“采样率”（Sample Rate）是指每秒钟采集多少个声音样本，或者说，是通过波形采样的方法记录 1 秒钟长度的声音需要多少个数据，其计量单位是赫兹（Hz）或千赫（kHz）。如果一种声音的采样率为 44kHz，那就表示每描述 1 秒钟的声音波形需要 44000 个数据。声卡的通用频率有三种：44.1kHz，22.05kHz，11.05kHz。原则上讲，采样率越高，采样点之间的间隔就越小，波形的描述就越精确，声音品质也就越好，但与之相应，其所需的存储空间也会越大。

## 2.3 量化级

“量化级”（Quantitative level）也称“量化数据位数”，是描述声音波形之数据位数的二进制数据，亦即每个采样点所能表示的数据范围，其通用单位是比特（bit）。标准 CD 音乐的量化级是 16bit，也就意味着 CD 记录声音的数据采用的是 16 位的二进制数。量化级是对模拟音频信号的幅度轴所进行的数字化，其取值大小直接决定着声音的动态范围，一般而言，量化级取值越高，所采信号动态范围就越大，数字化后的音频信号也就越接近数字化前的原始信号，其音质也越好，但数据量也就越大，所需的存储空间因而也越大。“采样率”和“量化级”是数字化声音的两个最基本的要素，是衡量数字化声音品质高低的重要指标。

## 2.4 声道数

声道（Sound Channels）是指声音在录制或播放时其音频信号所经过的空间通道，声道数即是声音录制时的音源数量，或声音回放时的扬声器数量。早期声卡多用单声道，缺乏立体方位感；后面出现了立体声，声音在录制过程中被分配到两股独立声道，从而较好地达到了声音定位效果，增强了聆听时的临场感受；此后又出现了四声道环绕技术，该技术规定了前左、前右，后左、后右这 4 个方位的发音点，听众被包围在中间，同时又增加了一个低音音箱，以加强对低频信号的回放处理，故也称 4.1 声道；再往后，杜比 AC-3 又在 4.1 声道的基础上增加了一个中置单元，用以负责传送低于 80Hz 的声音信号，这就是所谓的 5.1 声道；5.1 之后又有更强大的 7.1 系统问世，即在 5.1 的基础上又增加了中左和中右两个发音点，以求达到更加完美的境界，但由于成本比较高，尚未广泛普及。

## 2.5 信噪比

信噪比（Signal to Noise Ratio）又称“讯噪比”，是指特定参数（即“信/讯号”）值与非特异性参数（即“噪声”）的比值，是用有用信号功率（或电压）和噪声功率（或电压）比值的对数来表示的，计量单位多用分贝（dB）。信噪比越高表明杂音越少，声音回放时的质量也就越高，反之类推。信噪比一般不应该低于 70dB，高保真音箱的信噪比应达到 110dB 以上。

## 2.6 比特率

比特率 (Bit Rate) 是指每秒种所传输的比特 (bit) 数, 其常用单位是千比特每秒 (Bit Per Second, 简称 bps)。比特率越高, 表明单位时间内传输的数据量越大, 它指示了数字音频可以由播放器正常进行播放的速度值, 在网络环境中与宽带的流量也有关系, 网络的带宽必须足够高才能容纳多种内容的同时传输。音乐文件最常用的比特率是 128kbps, 例如, MP3 文件可以使用的一般在 8~320kbps 之间。比特率越高, 文件越大, 所占用的存储空间也相应越大。

## 2.7 杜比降噪

“杜比降噪” (Dolby Noise Reduction) 是指由英国物理学家瑞·弥尔顿·杜比 (Ray Milton Dolby, 1933-) 博士发明的一种降噪系统, 其目的在于降低录音或放音时伴随的噪声。1966 年发明了 DOLBY-A 型降噪系统, 主要用于专业机, 后来发明的 DOLBY-B 和 DOLBY-C 两套系统则主要用于家用设备。DOLBY-B 降噪系统为互补型降噪系统, 它通过提高音频信号高频段的信噪比来达到降噪目的, 该系统对高于 5kHz 的频率有大约 10dB 的降噪效果; DOLBY-C 降噪系统是由 DOLBY-B 型发展而来, 它对 1kHz 以上频率的降噪效果约为 20dB。

## 2.8 无损压缩

无损压缩 (Lossless Compression) 是利用数据的统计冗余进行压缩, 在不牺牲任何音频信号的前提下, 减少源 WAV 文件体积的压缩格式, 经过压缩的数据信号可完全回复原初状态而不引起任何失真, 因此也称“无失真压缩”或“可逆性压缩”。常见的无损压缩的音频格式为 APE、FLAC 等。

## 2.9 有损压缩

有损压缩 (Lossy Compression) 是利用人类对图像或声波中某些频率成分不够敏感的特性进行压缩, 在压缩过程中会有原始数据的损失且无法恢复, 损失的部分对理解原始图像或声音的影响较小, 却能换来较大的压缩比。常见的有损压缩的音频格式如 MP3, WAM, RA 等。

## 2.10 备案保存

出于存档备案的需要, 不仅要确保数字音频资源的数据完整、内容无误, 还要力求使该音频各方面的数据达到较高的技术指标, 以备长期保存和另派用场。

## 2.11 发布共享

出于服务共享的需要, 要求数字音频资源的数据完整、内容无误, 但数据各方面的技术指标并不要求太高, 以免使文件内存过大, 在发布共享时影响数据流量和使用效

果。

## 3 标准规范内容

### 3.1 制定规范的基本原则

本规范力求体现长效性、针对性和兼容性三个原则。所谓“长效性原则”是指，要尽量使数字化的音频具有长期可用、便于再加工和后期维护的优势；“针对性原则”则是要充分考虑不同类型的音频资源各自的特点、用途倾向、用户群以及数字化加工过程中的特殊性要求；所谓“兼容性原则”是指，要尽量使本规范与国内其他大型数字图书馆采用的标准规范乃至国际上通用的技术标准和操作规范能够兼容互通。

### 3.2 数字加工的操作流程

#### 3.2.1 资源整理

查看音频资源对象在物理形态和数据信息方面的完整性、正确性与可读性，并对之进行分类，可按信息内容将音频资料分为如下三类：一是以语言为核心表现手段的音频资料（简称“语言类”），具体包括演讲、朗诵、访谈、教学、新闻播报、故事演说（含评书）、话剧及体育类节目的录音等，其中的音频信息在经过识别后可转换成文本，这些资源虽可能包含音乐元素（如配乐诗朗诵，或具有音乐为背景的故事演播），但不以音乐为主；二是以音乐为核心表现手段的音频资料（简称“音乐类”），包括纯器乐曲、无伴奏的声乐曲，以及兼有声乐和器乐的各类乐曲资源，其中的音频信息在经过识别后可转换成乐谱（或可包含文字脚本），这些资源虽然有时包含有语言的因素（如歌剧或戏曲中的对白、宣叙与朗诵），但并不以语言为主；三是语言与音乐并重的音频资料（简称“语言—音乐类”），如说唱艺术、影视剧录音以及各种娱乐节目的录音等，若以书面形式记录这些资源信息，则其必需兼用文本和乐谱这两种符号体系。

#### 3.2.2 信息采集

具体又分为数字信号采集和模拟信号采集两种。数字信号采集：对于激光唱片（即CD）可采用音频工具软件直接抓取音轨信息，存放于计算机硬盘之上。对于数字录音带（DAT），则可先用相应的播放器播放，然后借助 SPDIF（SONY、PHILIPS 数字音频接口）数码界面输入计算机并存储到硬盘。模拟信号采集：对于胶质唱片[滚筒式，粗纹式(S.P.,78 转)，密纹式(L.P.,33 转)]、磁带等非数字信号，应根据音频资源各自的存储媒介，选择相应的播放设备作为模拟信号的输出端，在适宜的没有明显噪音和其他干扰因素的环境下进行播放。

### 3.2.3 格式选择

根据数字音频资源的两种应用级别：备案保存级和发布共享级别，分别设定不同的格式和技术标准。

#### 3.2.3.1 备案保存级

**WAV 格式：**经由微软公司（Microsoft）开发的一种声音文件格式（是 Wave[波形]一词的缩写），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，几乎被所有的音频处理软件和应用程序所广泛支持。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，因此其声音品质堪比 CD，也正因如此，其体积较大，占用的存储空间较多，不便于传输、发布和共享，但却是备案保存的理想格式之一。

**FLAC 格式：**其全称为 Free Lossless Audio Codec，意即无损耗的音频压缩编码，是专门针对 PCM 音频的特点设计的压缩方式。用这种方式生成的声音文件不会破坏原有的任何音频信息，可还原成 WAV 或 CD 的音质。FLAC 压缩文件可被播放器直接播放，而且是完全免费的，可被大多数操作系统所兼容，但其压缩比率不是很高，因而 FLAC 压缩文件的体积仍然较大，但不失为备案保存的有效格式。

**APE 格式：**也是一种无损的音频压缩方式，当把 APE 压缩文件还原为 WAV 格式后，音质不会有任何损害。与 FLAC 文件相比，APE 的压缩率较高，也就是说，同样大小的 WAV 文件在分别被转成 APE 格式和 FLAC 格式后，前者（APE）的体积更小，大约只有源文件（WAV 格式）的一半那么大，因而是既保真又轻便的理想化的压缩性音频格式。

**BWF 格式：**由欧洲广播联盟发展出来的一种广播波形格式，基于 RIFF-WAVE，可将数字音频与无线广播和电视进行无缝交换。作为一种特殊的 WAVE 文件，BWF 格式文件包含有音频内容的元数据信息，它支持时间戳与音轨标签，目前已被欧洲广播联盟（EBU）、美国保险会计与系统协会（IASA）、声频工程协会（AES）、澳大利亚档案馆（NAA）、美国国会图书馆（LC）、美国国家文理记录学院（NARAS）等知名机构设为音频保存的首要推荐格式。

资源对象	应用级别	参数指标			推荐格式
		采样率	量化级	声道数	
特殊录音资料	备案保存	≥96 kHz	24 bit	多声道、双声道或单声道，由原始资源对象自身的特性决定。	1.WAV- 无压缩或无损压缩； 2.APE- 无损压缩；
极高质量要求		≥48 kHz	24 bit		
CD 质量标准		44.1 kHz	≥16 bit		
录音最低质量		22.05	16 bit		

		kHz		3.FLAC-无损压缩
语音质量		22.05 kHz	16 bit	
语音最低质量		16 kHz	8 bit	

表 1：音频资料备案保存级加工标准

### 3.2.3.2 发布共享级

用于备案保存的数字音频资源在经过专业音频编辑软件的压缩转换后,可降低其某些技术指标,使其容量减小、便于传输上(下)载,以供发布共享之用。

**MP3 格式:** 其全称为 MPEG Audio Layer3,是一种诞生于 20 世纪 80 年代的有损性的音频压缩技术。相对于 WAV 文件而言,MP3 文件的压缩率可达 1:10 甚至 1:12,其优势在于能够在音质丢失很小的情况下把文件压缩到更小的程度,因此非常便于传输发布,在网络环境下应用极广。

**WMA 格式:** 其全称为 Windows Media Audio,也是微软公司推崇的一种音频格式,以减少数据流量的方法来达到更高的压缩比率,相对于 WAV 文件来说,WMA 文件的压缩率可达 1:18,比 MP3 还要便捷小巧,在源文件体积相同的情况下,压缩后的 WMA 文件就大小而言几乎是 MP3 文件的一半,可见其便捷性,但就音质而言却不如 MP3 文件好。

**AAC 格式:** 其全称为 Advanced Audio Coding,意为“高端的音频编码”,问世于 20 世纪末,是专为声音数据设计的文件压缩格式。AAC 文件虽也属于有损压缩格式,但相对于 MP3 文件而言,其体积更小,且音质更优,因而很适于网络环境下的发布与共享。

应用环境	应用级别	参数指标			推荐格式
		比特率/采样率	量化级	声道数	
比特率较高	发布共享	64~320kbps / 44.1kHz	16 bit	双声道	MP3 WMA
比特率较低		16~64KBPS / 22.05kHz		双/单声道	AAC

表 2：音频资料发布共享级加工标准

### 3.2.4 音频编辑

当数字音频信号是由模拟音频信号转换过来时，需要对采集的信号进行降噪处理、冗余信息处理、音量调整、均衡化处理等，此外，当数字资源在由“备案保存级”转换成“发布共享级”时，还需要压缩处理、转换格式或其他剪辑拼贴等工作。在以上两种编辑过程中，要确保音频资料有用信息的完整性，不留太多空白音频。

### 3.2.5 附件处理

附件在此特指音频资源自身附带的封面、封底、碟身信息、曲目注解（program notes）等物件，作为音频资料描述性元数据的载体，这些物件如能找到则需要扫描处理，发布共享级像素不低于 150ppi，格式不小于 JPEG，备案保存级推荐使用 300ppi 像素、Tiff 格式。数字化之后的音频信息及被处理后的附件信息应通过元数据相互关联最终组合成完整的便于快捷读取和听取的资源模块。

### 3.2.6 元数据处理

#### 3.2.6.1 元数据著录

无论是备案保存级还是发布共享级的音频资料，都需对其原始资源对象的元数据进行著录，描述性元数据著录推荐采用《大学数字图书馆国际合作计划》子项目成果《音频资料描述元数据标准规范》（附著录规则），管理性元数据推荐参照《国家图书馆管理元数据规范》的项目成果。

#### 3.2.6.1 元数据关联

在不同元数据之间、元数据与对象数据之间、元数据及其源载体数据之间实现关联，这是数据加工环节的重要步骤，关联方式参照《大学数字图书馆国际合作计划》子项目成果《音频资料描述元数据标准规范》中“相关资源”（Relation）条目的描述。

### 3.2.7 文件命名

数字音频文件的命名由以下三部分组成：加工单位识别码+资源类别代码+文献代码。各段代码之间无需连接符，直接顺次标记。其中的加工单位识别码由两位数字构成，由 CADAL 项目管理中心指定；资源类别码为一位字母，表示数字资源在信息内容方面的类别归属（参见附录 5.1）；文献代码为五位流水号，流水号不足五位数者在前面各位上以 0 补足。

### 3.2.8 成果保存

数字资源有两种保存机制：备案级和发布级，具体参数参见《国家数字图书馆数字资源长期保存规范》。在音频资源的数据信息时应同时保存其说明性的附件信息。

### 3.2.9 质量检测

根据数字化前资源对象的特殊性，以及数字化后的资源级别的不同，对元数据、元数据附件、存储媒介、保存机制等是否符合预设标准进行检测。在数据加工环节，需采用全面彻底的检验方式，内容包括：元数据验收、音频数据验收、附属物件验收、存储介质验收等；在数据验收环节，可采用抽检方式，抽检内容亦为：元数据验收、音频数据验收、附属物件验收、存储介质验收等。

## 4 参考文献

1. 王绍平：《科技部科技基础性工作专项资金重大项目》子课题成果《音频资料资源分析报告》（总项目组推荐稿），2003年7月。
2. 张成昱（等）：《我国数字图书馆标准规范建设》项目子课题成果《通用数字资源（音频数据）格式标准分析报告》，2003年8月。
3. 常林（等）：《我国数字图书馆标准规范建设》项目子课题成果《数字资源加工标准与操作指南》（子项目组定稿），2004年3月。
4. 聂华（等）：《图书馆专门资源对象的数字化加工标准推荐》，2004年5月。
5. 孙一钢（等）：《我国数字图书馆标准规范建设》项目子课题成果《数字资源加工标准》研究报告，2006年6月。
6. 孙一钢（等）：《我国数字图书馆标准规范建设》项目子课题成果《数字资源加工操作指南》研究报告，2006年6月。
7. 宋庆生：《国家图书馆对象数据项目》成果《音频数据加工标准与工作规范标准规范》（正式提交版），2010年6月。

## 5 附录

### 5.1 资源类别代码表

该代码由一位英文字母（详见表 3）构成；字母“L”（英语 Language 一词的首位字母）表示语言类音频资料，字母“M”（英语 Music 一词的首位字母）表示音乐类音频资料，字母“S”（英语 synthesis 的首位字母）表示语言-音乐相综合的音频资料。

类别代码	注解
L	语言类音频资料：以语言为核心表现手段的音频资料，具体包括演讲、朗诵、访谈、教学、新闻播报、故事演说（含评书）、话剧及体育类节目的录音等，其中的音频信息在经过识别后可转换成文本。
M	音乐类音频资料：以音乐为核心表现手段的音频资料，包括纯器乐曲、无伴奏的声乐曲，以及兼有声乐和器乐的各类乐曲资源，其中的音频信息在经过识别后可转换成乐谱。
S	语言-音乐综合类音频资料：语言与音乐并重的音频资料，如说唱艺术、影视剧录音以及各种娱乐节目的录音等，若以书面形式记录这些资源信息，则其必需兼用文本和乐谱这两种符号体系。

表 3：资源类别代码