Making Single File Audio CD Backups: How To Use Matroska to Do It
by Goldenear (edition 0.7 2004-08-01)

Introduction

What is our purpose?

Many people want to backup their original audio CD in order to be able to re-burn one of them in case of loss or degradation. The usual way to do it is to use a lossless encoded audio file with an associated cue file. You may also need several other files such as jpeg file(s) for the cover/artwork or text files for the lyrics. So, just for a single audio CD backup you have to manage several files, unless you put all of them in an archive file such as a zip, rar or tgz file. In any case, it's not really convenient.

Hopefully, there is an easy and convenient way to make a single file backup of an audio CD: you have to use the Matroska container format. This way, you will be able to keep the lossless audio, the CUE sheet information and all the other useful stuffs in a single file. And, ice on the cake, you'll be able to play a matroska file as simply as any audio file in player softwares such as Foobar2000 or mplayer. So, if you have enough hard-disk space, you can keep all the matroska files on your computer and have your entire music collection at your finger tip.

This Document explains how to create a matroska file backup of an audio CD and gives some technical information about this kind of file.

What is Matroska?

Matroska is a multimedia container format such as avi or quicktime/mpeg4. It should be able to handle any actual or future audio/video format/codec such as xvid, mpeg2, mp3 or Ogg Vorbis. Matroska can also handle different format of subtitle. There are several kinds of Matroska files, each of them with a specific extension:
– .mkv for audio/video files
– .mka for audio only files
– .mks for subtitles files
As we want here to use Matroska for an audio purpose we will use mka files.

For more information about Matroska and supported formats please consult the Matroska website at www.matroska.org.

Technical information

Organization of the data inside a Matroska file

On an audio CD you have one or more track. Each track has one or more index. Matroska has a similar but much more powerful way to organize the audio data:
A Matroska file has one or more segment which contain the other organizational elements. A segment can be seen as a global entity for a logical set (i.e. the set of all the tracks of a CD or the set of all the discs of a multi CD set). A segment can be organized in chapters. In this case, all the chapters are in an EditionEntry which is the top level of the chapters organization. Each chapter contains one or more sub-chapter and each sub-chapter contains one or more sub-chapter, etc... For more information about how a Matroska file is organized, please read the specs at:
http://www.matroska.org/technical/diagram/index.html

The Chapters

In the case of our backup purpose, an EditionEntry will refer to the global archived entity: If we backup a single audio CD, then EditionEntry will refer to the the disc itself and if we backup a multi CD set, then EditionEntry will refer to the whole CD set.

Inside an EditionEntry, there are the chapters. In the case of a single audio CD, the top level chapters correspond to the audio CD tracks and the first level of sub-chapters correspond to the indexes. In the case of a multi CD set, the top level chapters correspond to the discs of the CD set, the first level sub-chapters correspond to the tracks of each CD and the second level sub-chapters correspond to the indexes.

Different CD or different CD sets without any relationship between them should be placed into different segments and this usually means several mka files as there is usually only one segment per file.

As you can see, Matroska has a very powerful organizational system which can manage a whole album as well as a whole CD set in the same mka file.

The Tags

Track, EditionEntry, and Chapter/sub-chapters Elements can be tagged with useful information such as a title, an artist or a date. Matroska has a standardized tags naming scheme that can be found at:
http://www.matroska.org/technical/specs/tagging/index.html

For our backup purpose, we will use a hierarchical tags organization; that means that each tag will be affected to the entity corresponding to its level: In the case of a single audio CD, all the tags concerning to the whole album in its globality will be affected to the corresponding EditionEntry (e.g. Date or title of the album) and all the tags concerning to a track of the CD will be affected to the corresponding chapter (e.g. Title or performer of the song). In the case of a multi CD set, all the tags concerning to the whole CD set will be affected to the corresponding EditionEntry (e.g. Name or publisher of the CD set), all the tags concerning a whole disc of the set will be affected to the corresponding top level chapter and the tags concerning a track of the CD will be affected to the corresponding first level sub-chapter. If the mka file has no chapters, all the tags are affected to the track.

Here is a non exhaustive of the tags we will use:

In the case of a single audio CD backup:
    Tags for EntionEntry:
- TITLE: The global title of the CD/album. This is the same as using an ALBUM tag for each song, but in a more convenient centralized way.
- ARTIST: The name of the Artist for the whole CD; If the CD contains songs of different artists, we should have ARTIST="Various Artist"; In the case of a DJ mix, ARTIST can also be the name of the DJ.
- DATE: The date of the CD/album, generally, the publishing year.
- COMMENT: A comment applying to the entire CD/album.
- DISCID: The CDDB / FreeDB disc ID.
- CATALOG: The catalog number (sometimes the EAN/UPC).

    Tags for the top level chapters:
- TITLE: Title of the corresponding song/CD track.
- ARTIST: Artist/performer of the corresponding song/CD track.
- DATE: Date of the corresponding the song/CD track, generally the release year.

- GENRE: The genre of the corresponding song/CD track (e.g. Jazz, Rock, etc...).
- COMMENT: A comment applying to the corresponding song/CD track.
- SET_PART: This corresponds to the track number of the song/CD track.
- ISRC: International Standard Recording Number of the song/CD track.
- CD_TRACK_FLAGS: Some special flags of a CD track. This tag is only used to keep an information used to rebuild the CUE sheet.

In the case of a single audio CD backup:
Tags for EntionEntry:
- TITLE: The global title of the CD set (e.g. "Best of James Brown")
- ARTIST: The name of the Artist for the whole CD set.
- DATE: The date of the CD set, generally, the publishing year.
- COMMENT: A comment applying to the entire CD set.
- TOTAL_MEDIA_PART: A numerical value indicating the total number of media to this set. *We must always used this tag in the case of a multi CD set.*

Tags for the top level chapters:
- TITLE: The global title of the corresponding CD/album. This is the same as using an ALBUM tag for each song, but in a more convenient centralized way.
- ARTIST: The name of the Artist for the whole corresponding CD; If the CD contains songs of different artists, we should have ARTIST="Various Artist"; In the case of a DJ mix, ARTIST can also be the name of the DJ.
- DATE: The date of the corresponding CD/album, generally, the publishing year.
- COMMENT: A comment applying to the entire corresponding CD/album.
- DISCID: The CDDB / FreeDB disc ID.
- CATALOG: The catalog number (sometimes the EAN/UPC).

Tags for the first level sub-chapters:
- TITLE: Title of the corresponding song/CD track.
- ARTIST: Artist/performer of the corresponding song/CD track.
- DATE: Date of the corresponding the song/CD track, generally the release year.
- GENRE: The genre of the corresponding song/CD track (e.g. Jazz, Rock, etc...).
- COMMENT: A comment applying to the corresponding song/CD track.
- SET_PART: This corresponds to the track number of the song/CD track.
- ISRC: International Standard Recording Number of the song/CD track.
- CD_TRACK_FLAGS: Some special flags of a CD track. This tag is only used to keep an information used to rebuild the CUE sheet.

We can notice that we distinguish if a mka file is a backup of a singe audio CD or the backup of a whole CD set with the TOTAL_MEDIA_PART tag in the EditionEntry. This is very important.

On an audio CD, indexes are sometimes use to "separate" a song in different logical pieces: e.g. With symphonic music you can have different movements inside the same track and each movement can correspond to an index. In this case it should be useful to give a TITLE tag to the corresponding index/sub-chapter (e.g "first movement, piano").

For more information about tags, I recommend you to visit:
http://hobba.hobba.nl/audio/tag_frame_reference.html

And about cue <-> mka <-> foobar tags conversion, you may be interested in this:
http://www.pepper-prod.com/tags.html

How to do it

The general way to make a single mka file backup of an audio CD is very simple. First we need to extract the audio data from the CD using a ripping tool. This generally produce a big wav file of the whole CD. The ripping software should also produce a cue file and may automatically tag it with the help of CDDB/FreeDB. Then we have to encode the wave file using a (preferably) lossless codec; At the moment, Matroska supports FLAC or TTA. For more information about FLAC or TTA, please consult:
http://flac.sourceforge.net/ for information about FLAC.
http://www.true-audio.com/ for information about TTA.
We can also scan the cover/artwork and use the corresponding jpeg file(s). About scanning I recommend to do it at 300dpi as it won't produce too big jpeg files and it will provide a good enough quality for printing (if you need it later). You can also add text files of the lyrics (html files is fine) or an iso image file of the data track of a CD extra. We can then merge everything using a special Matroska tool, mkvmerge. This will produce the wanted single mka file. Now, let see how to do it in detail under Windows an under Linux.

Under Windows
//To be done soon!

Under Linux
//To be done soon!

References:

For Matroska:
http://www.matroska.org
http://www.matroska.org/technical/diagram/index.html
http://www.matroska.org/technical/specs/tagging/index.html
http://ld-anime.faireal.net/guide/jargon.matroska-en

For CD ripping softwares:
http://www.exactaudiocopy.de/ (Windows)
http://www.cdex.n3.net/ (Windows)
http://nostatic.org/grip/ (Linux)
http://www.xiph.org/paranoia/ (Linux)
http://cdrdao.sourceforge.net/ (Linux)

For useful tools:
http://www.bunkus.org/videotools/mkvtoolnix/index.html (Windows / Linux)
http://cuetools.sourceforge.net/ (Linux)

Notes:

Please report any error, typo, grammatical monstrosity or simply comment to goldenear at matroska.org.

Single mka file audio CD backup Howto
Copyright (c) 2004 Nicolas Le Guen
This text is under GNU FDL. For more information, please consult:
 http://www.gnu.org/copyleft/fdl.html

All logos, terms or trademarks are trademarks or registered trademarks of their respective owners.