Standards for Music Organization on Contemporary Filesystems

darkness

Version 0.3.1
Last revised: September 13, 2006
Generated: September 13, 2006

Contents

1 Introduction

This document specifies organizational standards for digital music stored on contemporary filesystems. The primary goal is to achieve consistency in organization and naming of music files.

This document relies heavily on "Standards for Music Information" [SMI], and uses many of the same terms defined in that document. Fields defined in [SMI] may be referred to as "SMI fields".

2 Conventions and assumptions

A *contemporary filesystem*, for purposes of this document, is a filesystem that contains files organized into hierarchies of directories. An example of contemporary filesystems would be VFAT (Microsoft Windows 95), NTFS (any version; Microsoft Windows NT, 2000), ext2 (Linux), HFS (MacOS), or UFS (many different "UNIX" operating systems). The filesystem should retain case, though it does not need to be case sensitive.

This document uses examples based on a typical UNIX filesystem. This means "`/`" will be used as a directory delimiter. For example, "`foo/bar`" indicates a directory named "`foo`", and a directory under that directory named "`bar`". This separator should be replaced with the appropriate delimiter for your particular filesystem.

Filesystem-specific nuances (for example, filename extensions to determine media type) are explicitly permitted. These nuances are not covered here unless otherwise noted, and any examples are free of any such nuances.

Each track is presumed to be stored in its own file. This standard is likely not applicable to other cases (for example, multiple tracks stored in a single file).

3 Directory structure

There is a directory structure that is used to group like tracks together. This directory structure first groups by artist, then by set and collection. This section describes the naming of both of these directories.

3.1 Artist grouping

The first level of grouping must be in a directory named after the contents of the SMI field artist name. For purposes of artist grouping, the artist name field must have a consistent name for all tracks in a set.

Compilation set, as defined in [SMI], are treated differently. Compilation albums must use methods 2 or 3 as defined in [SMI] section 4.2.1 for the artist name value when creating the artist grouping directory structure for a compilation set.

3.2 Set and collection grouping

The second level of grouping must be in the following format (references to SMI fields):

```
Set title (Set subtitle) - Collection title (Collection subtitle)
```

Any unset fields must be omitted, including their surrounding parentheses if any. The separating " - " is only allowed and required if there are one or more set (non-empty) fields on each side of the separator. There should be no leading or trailing spaces, and never two or more consecutive spaces, including in values from SMI fields. A sequence of two or more consecutive spaces must be replaced by a single space.

4 Filenames

Each track, after being placed under its proper directory structure, must be named in the following format:

```
TP. Track title (Track subtitle - Modification subtitle)
```

Where "TP" is the SMI field track position and the other portions are references directly to SMI field names. If either track subtitle or modification subtitle are unset, the separating hyphen (" - ") must be omitted. If both track subtitle and modification subtitle are unset, everything after the track title must be omitted. The track position must be given as a two-digit, zero padded number, followed by a period and exactly one space.

5 Character translations

SMI field values are to be left untouched whenever possible. The exception to this rule is character translations, which convert characters commonly illegal in filesystems to other character sequences (of zero or more characters). Characters in an SMI field value found in the first column of table 1 must be translated to those sequences found in the second column of that table.

5.1 Spaces

Two or more consecutive spaces within a file or directory name must be replaced with a single space. This conversion must be performed after character translations.

5.2 Other illegal characters

For filesystems which have illegal characters not listed in table 1, best judgment must be used. In most cases, the offending character should be deleted.

Table 1: Character translations

Illegal character	Translation
/	,[a]
\	,
:	,
*	Empty sequence[b]
?	Empty sequence
"	'
< (As mathematical operator)	`less than`
< (All other cases)	Empty sequence
> (As mathematical operator)	`greater than`
> (All other cases)	Empty sequence
|	Empty sequence

[a] A comma must be followed by exactly one space in this and all other translations.
[b] Effectively, "Empty sequence" means the illegal character must be deleted.

However, if removing the offending character would markedly impact the comprehension of the directory or file name in a negative way, a suitable replacement should be found.

6 Coalescing tags

When an SMI field is set but contains only tags, this field is considered to be a *tag-only SMI field*. When using an SMI field to construct a directory or file name, and one or more tag-only SMI fields immediately follow a set (non-empty) SMI field, the tags from the tag-only SMI fields must be added as new tags to the first preceeding non-tag-only SMI field. (See [SMI] for what it means to "add a tag.") After this merging, the tag-only SMI fields must be considered unset. This process is called *coalescing tags*.[1]

For example, take a track where the SMI set title is "`Exploding Sheep`", the SMI set subtitle is "`[GB]`", and the SMI collection subtitle is "`[CD 1]`". Without coalescing tags, the directory name for the second level of grouping would be "`Exploding Sheep ([GB]) - ([CD 1])`". After coalescing tags, the directory's name becomes "`Exploding Sheep [GB; CD1]`". Note that, since the tag-only fields are considered unset, the various punctuation that was previously added to accomodate them (such as the parentheses and the hyphen) is removed.

7 Revision history

```
Version 0.1:
- Initial version.
```

[1] Note that this concept is stolen practically wholesale from [SID3].

```
Version 0.2:
- Added rules for directory structure artist grouping in
  compilation albums.

Version 0.3:
- Added section on coalescing tags.

Version 0.3.1:
- Andy corrected my grammar: parenthesis->parentheses in many
  places.

$Revision: 1.4 $
```

References

[SMI] "Standards for Music Information." Vers. 0.3. 22 May 2002. 22 May 2002 <http://www.codefu.org/tagging/smi/smi.pdf>.

[SID3] "Standards for ID3 Tags." Vers. 0.5. 10 Apr 2002. 10 Apr 2002 <http://www.codefu.org/tagging/sid3/sid3.pdf>.