Object Linking and Embedding (OLE) Compound File (CF) format specification

Analysis of the OLECF format

Joachim Metz <joachim.metz@gmail.com>

Summary

OLECF is short for Object Linking and Embedding Compound File and is a container file format used for different purposes. OLECF is mainly known as the OLE2 File, Compound Binary File and Compound Document File format.

This document is intended as a working document for the OLECF specification. Which should allow existing Open Source forensic tooling to be able to process this file type.

Document information

Author(s): Joachim Metz <joachim.metz@gmail.com>

Abstract: This document contains information about the OLE Compound File (OLECF).

Classification: Public

Keywords: OLE, OLE2, Object Linking and Embedding, Compound Binary File, Compound Document File, Horrible Property File Format

License

Version

Version	Author	Date	Comments
0.0.1	J.B. Metz	December 2008 May 2009 June 2009	Initial version based on earlier notes.
0.0.2	J.B. Metz	January 2010	Added information about the compound object stream.
0.0.3	J.B. Metz	January 2010	Moved information about Word binary format to separate document.
0.0.4	J.B. Metz	December 2010	Minor changes.
0.0.5	J.B. Metz	July 2011	Minor changes.
0.0.6	J.B. Metz	May 2012	Minor changes.
0.0.7	J.B. Metz	September 2013	Added information about property set streams.
0.0.8	J.B. Metz	October 2013	Additional findings regarding directory entry creation time.
0.0.9	J.B. Metz	October 2013	Additional findings regarding non 512 sector size files and property set streams.

Table of Contents

1. Overview

The OLE (Object Linking and Embedding) CF (Compound File) format is used for different file types. It is mainly known by containing Microsoft Office 97 – 2003 documents. The file format is commonly known by the name OLE2 but also is referred to as:

- Compound Binary Format;
- Compound Document Format;
- Horrible Property File Format.

An OLE CF consist of the following distinguishable elements:

- file header
- data sectors (blocks) containing different types of information
  - Allocation Table
  - Directory
  - Stream

The OLE CF is very similar to the FAT file system.

- The different allocation tables contain information about the blocks used.
- The directory blocks contain the hierarchical relationship of the streams.
- The streams contain the content data.

Characteristics	**Description**
Byte order	Mainly little-endian, but can be set in file header
Date and time values	Filetime in UTC
Character string	ASCII strings are stored in extended ASCII with a codepage. Unicode strings are stored in UTF-16 little-endian without the byte order mark (BOM).

1.1. Test version

The following version of programs were used to test the information within this document:

- Microsoft Office
- OpenOffice

2. The file header

The file header is 512 bytes of size but note that the entire first sector is used to store it. The file header consists of:

offset	**size**	**value**	**description**
0	8		The signature (magic identifier) “\xd0\xcf\x11\xe0\xa1\xb1\x1a\xe1” “\x0e\x11\xfc\x0d\xd0\xcf\x11\x0e” The latter was used by older beta versions of OLE2 files
8	16		Class identifier (GUID)
24	2		Revision number of the file format (minor version)

offset	size	value	description
26	2		Version number of the file format (major version)
28	2		Byte order identifier \xff\xfe big endian \xfe\xff little endian
30	2		Size of a sector in the compound document file in power-of-two
32	2		Size of a short-sector (mini-sector) in the short-stream container stream in power-of-two
34	2	0	Reserved empty value
36	4	0	Reserved empty value
40	4	0	Reserved empty value
44	4		Total number of sectors used for the sector allocation table (SAT). The SAT is also referred to as the FAT (chain).
48	4		Sector identifier (SID) of first sector of the directory stream (chain).
52	4	0	Reserved empty value Used for transactioning
56	4		Minimum size of a standard stream (in bytes, most used size is 4096 bytes), streams smaller than this value are stored as short-streams
60	4		Sector identifier (SID) of first sector of the short-sector allocation table (SSAT). The SSAT is also referred to as Mini-FAT.
64	4		Total number of sectors used for the short-sector allocation table (SSAT).
68	4		Sector identifier (SID) of first sector of the master sector allocation table (MSAT). The MSAT is also referred to as Double Indirect FAT (DIF).
72	4		Total number of sectors used for the master sector allocation table (MSAT).
76	109 x 4 = 436		First part of the master sector allocation table (MSAT) containing 109 sector identifiers (SIDs).
512	...		Unknown (empty values)

2.1. File format version

Major version	Minor version	Description
3	33	
3	62	
4	62	Used by msninfo.dat

3. The allocation table

The OLE CF format contains multiple types of allocation tables:
- the Master Sector Allocation Table (MSAT);
- the Sector Allocation Table (SAT);
- the Short Sector Allocation Table (SSAT).

The allocation table contains an array of entries up to the sector size. An entry consists of:

offset	size	value	description
0	4		Sector identifier (SID)

Certain sector identifiers are used for the following purposes:

Sector identifier	Description
0xfffffffc (-4)	Marks the sector as used for the MSAT
0xfffffffd (-3)	Marks the sector as used for the SAT
0xfffffffe (-2)	Marks the sector as the end of the sector (FAT) chain
0xffffffff (-1)	Marks the sector as unused (free)

The actual file offset for a SID can be determined as following:

```
file offset = ( SID + 1 ) x sector size
```

The SID needs to be corrected by 1 to account for the first sector that contains the file header.

3.1. Master Sector Allocation Table (MSAT)

The MSAT start at offset 76 in the file header. The MSAT contains sector identifiers (SID) that refer to where SAT sector are situated in the OLE CF. In the file header there is room for 109 SID and last SID does not point to the next MSAT sector as in the MSAT sectors.

The MSAT does not contain a end of chain value, non used SIDs are marked as unused.

If the MSAT is larger than 109 SIDs additional MSAT sectors are used. The first MSAT sector can be found by the SID at offset 68 in the file header. It contains 0xfffffffe if the OLE CF does not contain any additional MSAT sectors. The MSAT sector contains similar data as the MSAT in the file header except that the last SID in the sector is used to point to the next MSAT sector. This value should be 0xfffffffe is the sector is the last sector in the MSAT chain.

The number of SIDs in a MSAT sector can be determined as following:

```
number of SIDs = ( sector size / 4 ) - 1
```

3.2. Sector Allocation Table (SAT)

The SAT sectors can be determined by the MSAT. The SAT contains chains of sector identifiers (SID) that refer to the next sector in the chain or the end of the chain. A SID chain contains the sequence of sectors in a stream.

The number of SIDs in a SAT sector can be determined as following:

```
number of SIDs = ( sector size / 4 )
```

3.3. Short Sector Allocation Table (SSAT)

The SSAT sectors can be determined by the MSAT. The SAT contains chains of sector identifiers (SID) that refer to the next sector in the chain or the end of the chain. A SID chain contains the sequence of sectors in a stream.

The number of SIDs in a SSAT sector can be determined as following:

```
number of SIDs = ( sector size / 4 )
```

4. The directory

The directory consists of multiple entries.

A directory entry is 128 bytes of size and consists of:

offset	size	value	description
0	64		The directory name in UTF-16 without the byte order mark (BOM) but with an end-of-string character
64	2		The byte size of the directory name including the end-of-string character a name of 4 characters has a size of 10 (4 + 1) x 2 = 10
66	1		The type of the directory entry See section: 4.1 Directory entry types
67	1		The node color of the directory entry. It refers to the color of the node in a red-black tree. 0x00 red 0x01 black
68	4		The directory identifier of the previous directory entry The value is -1 if no previous directory entry is present
72	4		The directory identifier of the next directory

offset	size	value	description
			entry The value is -1 if no next directory entry is present
76	4		The directory identifier of the child directory entry The value is -1 if no child directory entry is present
80	16		Class identifier
96	4		User flags
100	8		Creation time Contains a Windows Filetime
108	8		Modification time Contains a Windows Filetime
116	4		Sector identifier (SID) of the first sector of the directory Refers to the SID of a stream or the SID of short-stream container stream
120	4		The byte size of the directory Refers to the size of a stream or the size of a short-stream container stream
124	4		Reserved

Note that in Office template OLECF files the creation time can be set to -1 (0xffffffffffffffff). It is unknown what this value should represent. This behavior has not been observed for the modification time.

Is there a difference in how empty directory entries are stored?

```
00000000: 00 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00   ........ ........
00000010: 00 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00   ........ ........
00000020: 00 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00   ........ ........
00000030: 00 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00   ........ ........
00000040: 00 00 00 00 ff ff ff ff  ff ff ff ff ff ff ff ff   ........ ........
00000050: 00 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00   ........ ........
00000060: 00 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00   ........ ........
00000070: 00 00 00 00 fe ff ff ff  00 00 00 00 00 00 00 00   ........ ........
```

Note some OLECF files do not have directory entries at all.

4.1. Directory entry types

Value	Identifier	Description
0x00		empty
0x01		storage
0x02		stream

Value	Identifier	Description
0x03		lock bytes
0x04		property
0x05		root storage

4.2. Root directory

A directory with the name “RootEntry” should be the first directory entry in the directory chain (stream). The root directory should have a type of 0x05. Some older implementations of OLE CF only store the name “R”.

4.3. Short-sector streams

When the size of a stream is smaller than the minimum size of a standard stream the data of that stream is stored in the short-sector stream. The size and the sector identifier of the short-sector stream is stored in the root directory.

TODO: describe how to determine the file offset of a SSAT identifier (SSID)?

The actual file offset for a SSID can be determined as following:

```
file offset = ( SSID x short-sector size ) + offset short-sector stream
```

4.4. Shared property streams

All shared property sets are identified by a stream or storage name with the prefix "\005" (or 0x05) to show that it is a property set that can be shared among applications. The Summary Information property set is no exception. The name of the stream that contains the Summary Information property set is: "\005SummaryInformation"

5. Compound object stream

The object stream contains binary data for embedded objects. Word has no knowledge of the contents of this stream.

The compound object specifies the Clipboard Format and the display name of the linked object or embedded object. The compound object stream has the name "\1CompObj". A compound object consists of:

- the compound object header
- the compound object stream

Characteristics	Description
Byte order	Dependent on the byte order in the OLECF file header?
Character string	ASCII strings are stored in extended ASCII with a codepage. Unicode strings are stored in UTF-16 little-endian without the byte order mark (BOM).

5.1. The compound object header

The compound object header 28 bytes of size and consists of:

offset	size	value	description
0	4		Unknown Reserved 16-bit 1 followed by 0xfe 0xff (little-endian)
4	4		Unknown Version
8	20		Unknown Reserved 32-bit -1 followed by a GUID

5.2. The compound object stream

The compound object stream is variable of size and consists of:

offset	size	value	description
0	4		User type string size
4	...		User type string extended ASCII string terminated by an end-of-string character
...	4		Clipboard format data size 0xfffffffe => 4 bytes standard clipboard format identifier 0xffffffff => 4 bytes standard clipboard format identifier 1 to 0x190 => registered clipboard format identifier data size 0 => no data present
...	...		Clipboard format data A 4 byte standard clipboard format identifier or an extended ASCII string terminated by an end-of-string character contain the the name of a registered clipboard format identifier.
...	4		Reserved string size If this value is 0 or > 0x28 the remainder of the compound object stream should be ignored
...	...		Reserved string extended ASCII string terminated by an end-of-string character
...	4		Unicode marker If this value is not 0x71b239f4 the remainder of the compound object stream should be ignored
...	4		Unicode user type string size

offset	size	value	description
...	...		Unicode user type string UTF-16 little-endian string terminated by an end-of-string character
...	4		Unicode clipboard format data size 0xfffffffe => 4 bytes standard clipboard format identifier 0xffffffff => 4 bytes standard clipboard format identifier 1 to 0xfffffffd => registered clipboard format identifier data size 0 => no data present
...	...		Unicode clipboard format data A 4 byte standard clipboard format identifier or an extended ASCII string terminated by an end-of-string character contain the the name of a registered clipboard format identifier.
...	4		Unicode reserved string size
...	...		Unicode reserved string UTF-16 little-endian string terminated by an end-of-string character

5.2.1. Standard clipboard format identifiers

Value	Identifier	Description
0x00000002	CF_BITMAP	Bitmap16 Object structure (BMP)
0x00000003	CF_METAFILEPICT	Windows Metafile (WMF)
0x00000008	CF_DIB	Device Independent Bitmap Object structure (WMF DIB)
0x0000000e	CF_ENHMETAFILE	Enhanced Metafile (EMF)

5.2.2. Registered clipboard format identifiers

TODO

6. Property set streams

Certain streams are property set streams. These streams contain information defined as properties. A property set stream consists of:

- the property set header
- the property set section list
- multiple the property set sections consisting of

- the property set section header
- the property set properties list
- multiple property set properties

Characteristics	Description
Byte order	Dependent on the byte order value in the property set header
Character string	ASCII strings are stored in extended ASCII with a codepage. Unicode strings are stored in UTF-16 little-endian without the byte order mark (BOM).

6.1. The property set header

The property set header is 28 bytes of size and consists of:

offset	size	value	description
0	2		The byte order \xff\xfe big endian \xfe\xff little endian
2	2		The format
4	4		The system version The upper 16-bit contain the operating platform type 0x0000 for Win16 0x0001 for Macintosh 0x0002 for Win32
8	16		The class identifier
24	4		The number of sections in the stream

6.2. The property set section list

The property set header is followed by the property set section list entries. A property set section list entry is 20 bytes of size and consists of:

offset	size	value	description
0	16		The class identifier
16	4		The offset relative from the start of the property set header

6.2.1. The property set class identifier

Class identifier	Description
6444048f-4c8b-11d1-8b70-080036b11a03	Image summary information (FMTID_ImageSummaryInformation)
64440490-4c8b-11d1-8b70-080036b11a03	Audio summary information (FMTID_Audio, FMTID_AudioSummaryInformation)

Class identifier	Description
64440491-4c8b-11d1-8b70-080036b11a03	Video (FMTID_Video)
64440492-4c8b-11d1-8b70-080036b11a03	Media file (FMTID_MediaFile)
f29f85e0-4ff9-1068-ab91-08002b27b3d9	Summary information (FMTID_SummaryInformation)
d5cdd502-2e9c-101b-9397-08002b2cf9ae	Document summary information (FMTID_DocSummaryInformation)
d5cdd505-2e9c-101b-9397-08002b2cf9ae	User defined (FMTID_UserDefinedProperties)
000214A1-0000-0000-C000-000000000046	Internet site (FMTID_InternetSite)
56A3372E-CE9C-11D2-9F0E-006097C686F6	Music (FMTID_Music)

6.3. The property set section header

A property set section header is 8 bytes of size and consists of:.

offset	size	value	description
0	4		The properties data size
4	4		The number of properties in the section

6.4. The property set section property list

The property set section header is followed by the property set section property list entries. A property set section property list entry is 8 bytes of size and consists of:

offset	size	value	description
0	4		The property identifier
4	4		The property data offset The offset is relative from the start of the property set section header

6.5. The property set section property

The property set section property list is followed by the property set section properties. A property set section property is variable of size and consists of:

offset	size	value	description
0	4		The property type
4	4		The property data dependent on the type

6.6. The property types

See document containing OLE property type definition.

The first 4 bytes of the data of variable sized property type contains the size of the data that follows the 4 bytes.

Note that strings can be stored with additional 0-byte values and are likely stored 32-bit aligned.

6.7. The property identifiers

Summary Information properties

Note that for the property identifiers the variants PIDSI_TITLE and PID_TITLE are used interchangeably. Other known variants are GKPIDSI_TITLE.

Value	Type	Property identifier	Description
0x0000		PIDSI_DICTIONARY	Reserved
0x0001	0x0002	PIDSI_CODEPAGE	Codepage Signed 16-bit Integer The codepage of the strings in the property section
0x0002	0x001e 0x001f	PIDSI_TITLE	Title ASCII or Unicode string
0x0003	0x001e 0x001f	PIDSI_SUBJECT	Subject ASCII or Unicode string
0x0004	0x001e 0x001f	PIDSI_AUTHOR	Author ASCII or Unicode string
0x0005	0x001e 0x001f	PIDSI_KEYWORDS	Keywords ASCII or Unicode string
0x0006	0x001e 0x001f	PIDSI_COMMENTS	Comments ASCII or Unicode string
0x0007	0x001e 0x001f	PIDSI_TEMPLATE	Template ASCII or Unicode string
0x0008	0x001e 0x001f	PIDSI_LASTAUTHOR	Last Saved By ASCII or Unicode string
0x0009	0x001e 0x001f	PIDSI_REVNUMBER	Revision Number ASCII or Unicode string
0x000a	0x0040	PIDSI_EDITTIME	Total editing time Filetime Value contains a duration
0x000b	0x0040	PIDSI_LASTPRINTED	Last printed date and time Filetime

Value	Type	Property identifier	Description
0x000c	0x0040	PIDSI_CREATE_DTM	Creation date and time Filetime
0x000d	0x0040	PIDSI_LASTSAVE_DTM	Last saved date and time Filetime
0x000e	0x0003	PIDSI_PAGECOUNT	Number of pages Signed 32-bit Integer
0x000f	0x0003	PIDSI_WORDCOUNT	Number of words Signed 32-bit Integer
0x0010	0x0003	PIDSI_CHARCOUNT	Number of characters Signed 32-bit Integer
0x0011	0x0047	PIDSI_THUMBNAIL	Thumbnail TODO
0x0012	0x001e 0x001f	PIDSI_APPNAME	Creating application name ASCII or Unicode string
0x0013	0x0003	PIDSI_SECURITY	Security Signed 32-bit Integer TODO add reference to section below

Security

Value	Identifier	Description
0x00000000	SECURITY_NONE	None No security states specified by the property
0x00000001	SECURITY_PASSWORD	Password protected The document MUST be password protected
0x00000002	SECURITY_READONLYRECOMMEND	Read-only recommended The document read-only is recommended but not enforced
0x00000004	SECURITY_READONLYENFORCED	Read-only enforced The document is always opened read-only
0x00000008	SECURITY_LOCKED	Locked for annotations The document is always opened read-only except for annotations

Document Summary Information properties

Note that for the property identifiers the variants PIDDSI_CATEGORY and PID_CATEGORY are used interchangeably. Other known variants are GKPIDDSI_CATEGORY.

Value	Type	Property identifier	Description
0x0001	0x0002	PIDDSI_CODEPAGE	Codepage The codepage of the strings in the property

Value	Type	Property identifier	Description
			section
0x0002	0x001e 0x001f	PIDDSI_CATEGORY	Category ASCII or Unicode string
0x0003	0x001e 0x001f	PIDDSI_PRESFORMAT	Presentation format ASCII or Unicode string TODO add reference to section below
0x0004	0x0003	PIDDSI_BYTECOUNT	Number of bytes (in document) Signed 32-bit Integer
0x0005	0x0003	PIDDSI_LINECOUNT	Number of lines (in document) Signed 32-bit Integer
0x0006	0x0003	PIDDSI_PARCOUNT	Number of paragraphs (in document) Signed 32-bit Integer
0x0007	0x0003	PIDDSI_SLIDECOUNT	Number of slides (in document) Signed 32-bit Integer
0x0008	0x0003	PIDDSI_NOTECOUNT	Number of notes (in document) Signed 32-bit Integer
0x0009	0x0003	PIDDSI_HIDDENCOUNT	Number of hidden slides (in document) Signed 32-bit Integer
0x000a	0x0003	PIDDSI_MMCLIPCOUNT	Number of multimedia clips (in document) Signed 32-bit Integer
0x000b	0x000b	PIDDSI_SCALE	Scale Boolean
0x000c	0x100c	PIDDSI_HEADINGPAIR	Heading pair Vector of Variant values
0x000d	0x101e 0x101f	PIDDSI_DOCPARTS	Document parts Vector of ASCII string values
0x000e	0x001e 0x001f	PIDDSI_MANAGER	Manager ASCII or Unicode string
0x000f	0x001e 0x001f	PIDDSI_COMPANY	Company ASCII or Unicode string
0x0010	0x000b	PIDDSI_LINKSDIRTY	Links dirty Boolean True if the links have changed outside the application.
0x0011	0x0003	PIDDSI_CCHWITHSPACES	Number of characters including white-space (in document)
0x0013	0x000b	PIDDSI_SHAREDDOC	Shared document Boolean According to MSDN this always must be false.
0x0014		PIDDSI_LINKBASE	Link base According to MSDN this value must not be

Value	Type	Property identifier	Description
			written
0x0015		PIDDSI_HLINKS	Hyper links According to MSDN this value must not be written
0x0016	0x000b	PIDDSI_HYPERLINKSCHANGED	Hyper links changed Boolean True if the hyper links have changed outside the application.
0x0017	0x0003	PIDDSI_VERSION	Creating application version Signed 32-bit Integer Where the major version is stored in the upper 16-bit and the minor version in the lower 16-bit. E.g. a value of 0x000e0000 represents 14.0
0x0018	0x0041	PIDDSI_DIGSIG	Digital signature Binary data (BLOB)
0x001a	0x001e 0x001f	PIDDSI_CONTENTTYPE	Content type ASCII or Unicode string
0x001b	0x001e 0x001f	PIDDSI_CONTENTSTATUS	Content status ASCII or Unicode string
0x001c	0x001e 0x001f	PIDDSI_LANGUAGE	Language ASCII or Unicode string
0x001d	0x001e 0x001f	PIDDSI_DOCVERSION	Document version ASCII or Unicode string

TODO
GKPIDDSI_HEADINGPAIR
0x0000000C
MUST be a VtVecHeadingPair property (section 2.3.3.1.15). Each VtHeadingPair element (section 2.3.3.1.13) in VtVecHeadingPair.vtValue.rgHeadingPairs defines a heading string and a count of document parts as found in the GKPIDDSI_DOCPARTS property (section 2.3.3.2.2.1) to which this heading applies. The total sum of document counts for all headers in this property MUST be equal to the number of elements in the GKPIDDSI_DOCPARTS property (section 2.3.3.2.2.1) property.

GKPIDDSI_DOCPARTS
0x0000000D
MUST be a VtVecUnalignedLpstr (section 2.3.3.1.10) or VtVecLpwstr property (section 2.3.3.1.8). Each string element of the vector specifies a part of the document. The elements of this vector are ordered according to the header they belong to as defined in the GKPIDDSI_HEADINGPAIR property (section 2.3.3.2.2.1).
Example: The first element of the heading pair vector indicates that it has four document parts associated with it. Elements 1 to 4 of the document parts vector are grouped under this header. The next element of the heading pair vector indicates that it has three document parts associated with it. The document part vector elements 5 to 7 are grouped under this header, and so on.

GKPIDDSI_DIGSIG
0x00000018
MUST be a VtDigSig property (section 2.3.3.1.17). VtDigSig.vtValue specifies the data of the VBAdigital signature (2) for the VBA project embedded in the document. MUST NOT exist if the VBA project of the document does not have a digital signature (2) or if the project is absent. MAY be ignored.<49>

Presentation format

Value	Description
	None (Empty string)
On-screen Show	
On-screen Show (4:3)	
Letter Paper (8.5x11 in)	
Ledger Paper (11x17 in)	
A3 Paper (297x420 mm)	
A4 Paper (210x297 mm)	
B4 (ISO) Paper (250x353 mm)	
B5 (ISO) Paper (176x250 mm)	
B4 (JIS) Paper (257x364 mm)	
B5 (JIS) Paper (182x257 mm)	
Hagaki Card (100x148 mm)	
35mm Slides	
Overhead	
Banner	
Custom	
On-screen Show (16:9)	
On-screen Show (16:10)	

Image Summary Information properties

Value	Type	Property identifier	Description
0x0002	0x001f	PIDISI_FILETYPE	File type
0x0003	0x0013	PIDISI_CX	Width
0x0004	0x0013	PIDISI_CY	Height
0x0005	0x0013	PIDISI_RESOLUTIONX	Horizontal resolution
0x0006	0x0013	PIDISI_RESOLUTIONY	Vertical resolution

Value	Type	Property identifier	Description
0x0007	0x0013	PIDISI_BITDEPTH	Bit depth
0x0008	0x001f	PIDISI_COLORSPACE	Color space
0x0009	0x001f	PIDISI_COMPRESSION	Compression
0x000a	0x0013	PIDISI_TRANSPARENCY	Transparency
0x000b	0x0013	PIDISI_GAMMAVALUE	Gamma value
0x000c	0x0013	PIDISI_FRAMECOUNT	Frame count
0x000d	0x001f	PIDISI_DIMENSIONS	Dimensions

Music properties

Value	Type	Property identifier	Description
0x0002		PIDSI_ARTIST	Artist
0x0003		PIDSI_SONGTITLE	Song title
0x0004		PIDSI_ALBUM	Album
0x0005		PIDSI_YEAR	Year
0x0006		PIDSI_COMMENT	Comment
0x0007		PIDSI_TRACK	Track
0x000b		PIDSI_GENRE	Genre
0x000c		PIDSI_LYRICS	Lyrics

Video properties

Value	Type	Property identifier	Description
0x0002	0x001f	PIDVSI_STREAM_NAME	Stream name
0x0003	0x0013	PIDVSI_FRAME_WIDTH	Frame width
0x0004	0x0013	PIDVSI_FRAME_HEIGHT	Frame height
0x0005	0x0013	PIDVSI_FRAME_C	Number of frames

Value	Type	Property identifier	Description
		OUNT	
0x0006	0x0013	PIDVSI_FRAME_RATE	Frame rate Value in frames per milliseconds
0x0007	0x0013	PIDVSI_TIMELENGTH	Time length Value in milliseconds
0x0008	0x0013		Data rate Value in bytes per milliseconds
0x0009	0x0013		Sample size
0x000a	0x001f		Compression
0x000b	0x0012		Stream number

Audio summary information properties

Value	Type	Property identifier	Description
0x0002	0x0008	PIDASI_FORMAT	
0x0003	0x0013	PIDASI_TIMELENGTH	Value in milliseconds
0x0004	0x0013	PIDASI_AVG_DATA_RATE	Value in Hz
0x0005	0x0013	PIDASI_SAMPLE_RATE	Value in bits
0x0006	0x0013	PIDASI_SAMPLE_SIZE	Value in bits
0x0007	0x0013	PIDASI_CHANNEL_COUNT	
0x0008	0x0012	PIDASI_STREAM_NUMBER	
0x0009	0x001f	PIDASI_STREAM_NAME	
0x000a	0x001f	PIDASI_COMPRESSION	

7. Notes

There are multiple type of data sectors

- MSAT sector (marked by 0xfffffffd (-3) in the SAT)
  Consist of (sector size / 4) MSAT sector values
- SAT sector (marked by 0xfffffffc (-4) in the SAT)
  Consist of (sector size / 4) SAT sector values
- directory sector

Consists of (sector size / short-sector size) directory entries

- empty sector (marked by 0xffffffff (-1) in the SAT)
- other sector (marked by a positive value in the SAT)

an 0xfffffffe (-2) in the SAT marks end of chain

a directory entry consists of:

msinfo.exe

http://msdn.microsoft.com/en-us/library/windows/desktop/aa370310(v=vs.85).aspx

References for DocumentSummaryInformation stream:

http://msdn.microsoft.com/en-us/library/dd945671%28v=office.12%29.aspx

http://msdn.microsoft.com/en-us/library/windows/desktop/aa380374%28v=vs.85%29.aspx

https://github.com/alexbevi/redmine_msg_preview/blob/master/data/FileInfo.pas

Appendix A. References

Title: DIG2000 file format proposal – Appendix A
Author(s): Digital Imaging Group
URL: http://www.i3a.org/pdf/wg1n1017.pdf
Date: October 30, 1998

Title: OpenOffice - Microsoft Compound Document File Format
Author(s): Daniel Rentz
URL: http://sc.openoffice.org/compdocfileformat.pdf
Date: August 30, 2004

Title: Advanced Authoring Format (AAF) Low-Level Container Specification v1.0.1
Author(s): AAF Association
URL: http://sourceforge.net/projects/aaf/
Date: 2004

Title: Windows Compound Binary File Format Specification
Author(s): Microsoft
URL: http://download.microsoft.com/download/0/B/E/0BE8BDD7-E5E8-422A-ABFD-4342ED7AD886/WindowsCompoundBinaryFileFormatSpecification.pdf
Date: 2007

[MSDN]
Title: Microsoft Developer Network
URL: http://msdn.microsoft.com/

Title: The Summary Information Property Set
URL: http://msdn.microsoft.com/en-us/library/aa380376(VS.85).aspx
URL: http://msdn.microsoft.com/en-us/library/windows/desktop/aa380376%28v=vs.85%29.aspx

Title: Summary Information Stream Property Set
URL: http://msdn.microsoft.com/en-us/library/aa372045.aspx

Title: SummaryInformation
URL: http://msdn.microsoft.com/en-us/library/dd942545.aspx

Title: PIDSI
URL: http://msdn.microsoft.com/en-us/library/dd925819%28v=office.12%29.aspx

Title: PIDDSI
URL: http://msdn.microsoft.com/en-us/library/dd945671%28v=office.12%29.aspx

[MS-OLEPS]
Title: [MS-OLEPS] Object Linking and Embedding (OLE) Property Set Data Structures
URL: http://msdn.microsoft.com/
Date: August 12, 2009

[MS-OLEDS]
Title: [MS-OLEDS] Object Linking and Embedding (OLE) Data Structures Structures

URL: http://msdn.microsoft.com/
Date: December 18, 2009

Appendix B. GNU Free Documentation License

Version 1.3, 3 November 2008

```
Copyright © 2000, 2001, 2002, 2007, 2008 Free Software Foundation, Inc.
<http://fsf.org/>

Everyone is permitted to copy and distribute verbatim copies of this license
document, but changing it is not allowed.
```

0. PREAMBLE
The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS
This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a

section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

The "publisher" means any person or entity that distributes copies of the Document to the public.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- D. Preserve all the copyright notices of the Document.
- E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

- G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- H. Include an unaltered copy of this License.
- I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS
You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS
A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION
Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION
You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and

will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE
The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING
"Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

"Incorporate" means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the

same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.