

## Exiv2 - Feature #1108

### Recursively dump sub-files of an image

21 Aug 2015 08:11 - Robin Mills

<b>Status:</b>	Closed	<b>Start date:</b>	21 Aug 2015
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Robin Mills	<b>% Done:</b>	100%
<b>Category:</b>	metadata	<b>Estimated time:</b>	35.00 hours
<b>Target version:</b>	0.26		

#### Description

In the discussion of [#1105](#), I've a discovery about sub-files, thanks to a comment by Jeroen in this thread: <http://dev.exiv2.org/boards/3/topics/1131>

The "improper" data in the Sony1 image is the preview. Here's the proof:

#### 1 Dump the Structure of DSC01825.jpg

```
561 rmills@rmillssmbp:~/temp/foo $ exiv2 -pS DSC01825.jpg
STRUCTURE OF JPEG FILE: DSC01825.jpg
address | marker | length | data
  2 | 0xd8 SOI | 0
  4 | 0xe1 APP1 | 48842 | Exif..II*.....
48848 | 0xe2 APP2 | 304 | MPF.II*.....0100.....
49154 | 0xdb DQT | 132
49288 | 0xc4 DHT | 418
49708 | 0xc0 SOF0 | 17
49727 | 0xda SOS | 12
```

\$

#### 2 Extract the APP1 segment into buff.tif and dump that:

```
562 rmills@rmillssmbp:~/temp/foo $ dd bs=1 skip=12 count=48842 if=DSC01825.jpg of=buff.tif ; exiv2
-pS buff.tif
48842+0 records in
48842+0 records out
48842 bytes transferred in 0.125983 secs (387687 bytes/sec)
STRUCTURE OF TIFF FILE (II): buff.tif
address | tag | type | count | offset | value
  10 | 0x010e ImageDescription | ASCII | 32 | 158 |
  22 | 0x010f Make | ASCII | 5 | 190 | SONY
  34 | 0x0110 Model | ASCII | 8 | 196 | ILCE-7R
  46 | 0x0112 Orientation | SHORT | 1 | 1 | 1
  58 | 0x011a XResolution | RATIONAL | 1 | 204 | 204/0
  70 | 0x011b YResolution | RATIONAL | 1 | 212 | 212/0
  82 | 0x0128 ResolutionUnit | SHORT | 1 | 2 | 2
  94 | 0x0131 Software | ASCII | 14 | 220 | ILCE-7R v2.00
 106 | 0x0132 DateTime | ASCII | 20 | 234 | 2015:07:09 00:47:5
6
 118 | 0x0213 YCbCrPositioning | SHORT | 1 | 2 | 2
 130 | 0x8769 ExifTag | LONG | 1 | 360 | 360
 142 | 0xc4a5 PrintImageMatching | UNDEFINED | 106 | 254 | ...
38170 | 0x0103 Compression | SHORT | 1 | 6 | 6
38182 | 0x010e ImageDescription | ASCII | 32 | 38330 |
38194 | 0x010f Make | ASCII | 5 | 38362 | SONY
38206 | 0x0110 Model | ASCII | 8 | 38368 | ILCE-7R
38218 | 0x0112 Orientation | SHORT | 1 | 1 | 1
38230 | 0x011a XResolution | RATIONAL | 1 | 38376 | 38376/0
38242 | 0x011b YResolution | RATIONAL | 1 | 38384 | 38384/0
38254 | 0x0128 ResolutionUnit | SHORT | 1 | 2 | 2
38266 | 0x0131 Software | ASCII | 14 | 38392 | ILCE-7R v2.00
```

38278		0x0132	DateTime		ASCII		20		38406		2015:07:09 00:47:5
6											
38290		0x0201	JPEGInterchangeFormat		LONG		1		38426		38426
38302		0x0202	JPEGInterchangeFormatLeng		LONG		1		10408		10408
38314		0x0213	YCbCrPositioning		SHORT		1		2		2
\$											

3 Extract the 0x0201 JPEGInterchangeFormat record (of length JPEGInterchangeFormatLength) into buff.jpg and dump that:

```
563 rmills@rmillssmbp:~/temp/foo $ dd bs=1 skip=38426 count=10408 if=buff.tif of=buff.jpg ; exiv2 -
pS buff.jpg
10408+0 records in
10408+0 records out
10408 bytes transferred in 0.045672 secs (227886 bytes/sec)
STRUCTURE OF JPEG FILE: buff.jpg
address | marker | length | data
  2 | 0xd8 SOI | 0
  4 | 0xdb DQT | 132
 138 | 0xc4 DHT | 418
 558 | 0xc0 SOF0 | 17
 577 | 0xda SOS | 12
```

It's a valid little jpg. Open it. It's the preview.

4 Now let's examine the previews:

```
576 rmills@rmillssmbp:~/temp/foo $ exiv2 -pp DSC01825.jpg
Error: Offset of directory Sony1, entry 0x2001 is out of bounds: Offset = 0x00901076; truncating t
he entry
Preview 1: image/jpeg, 160x120 pixels, 10408 bytes
577 rmills@rmillssmbp:~/temp/foo $ 571 rmills@rmillssmbp:~/temp/foo $ exiv2 --verbose --force -epl
DSC01825.jpg
File 1/1: DSC01825.jpg
Error: Offset of directory Sony1, entry 0x2001 is out of bounds: Offset = 0x00901076; truncating t
he entry
Writing preview 1 (image/jpeg, 160x120 pixels, 10408 bytes) to file ./DSC01825-preview1.jpg
```

5 DSC01825-preview1.jpg and buff.jpg are identical.

```
567 rmills@rmillssmbp:~/temp/foo $ 572 rmills@rmillssmbp:~/temp/foo $ ls -alt *.jpg
-rw-r--r--+ 1 rmills staff 10408 14 Aug 11:44 DSC01825-preview1.jpg
-rw-r--r--@ 1 rmills staff 10408 14 Aug 11:31 buff.jpg
-rw-r--r--@ 1 rmills staff 10125312 13 Aug 17:52 DSC01825.jpg
573 rmills@rmillssmbp:~/temp/foo $ diff buff.jpg DSC01825-preview1.jpg
574 rmills@rmillssmbp:~/temp/foo $ md5 buff.jpg DSC01825-preview1.jpg
MD5 (buff.jpg) = 4d49a9ce3d980b69bfa129e05483b041
MD5 (DSC01825-preview1.jpg) = 4d49a9ce3d980b69bfa129e05483b041
575 rmills@rmillssmbp:~/temp/foo $
```

I'm delighted by this discovery because I've been contemplating buying a Sony Alpha 7 Mirrorless camera and it seems that exiv2 is going to complain about 0x0201 with every image. Not good.

I think it would be better for libexiv2 to respect this situation and suppress the warning. We should however validate the integrity of the situation before deciding to suppress. It's very likely that some software could rewrite the image and blindly copy/relocate the APP1 segment with the resulting 0x0201 address being wrong. Exiv2 should report that situation.

Another reason for being pleased about this discovery is to restore my admiration of Sony. Yesterday I was thinking "Why have Sony not fixed a bug that has been in their Exif firmware for at least 5 years?". Now I'm not so sure it's a bug. The preview is embedded in the APP1 segment.

And the final reason for being pleased is to see how effectively the -pS option can be used to analyse this file.

This feature request is to add option -pR to exiv2 to recursively dump subfiles in the image.

#### Related issues:

Related to Exiv2 - Bug #1105: exiv2 output is inconsistent and seemingly rand...	Closed	13 Aug 2015
Related to Exiv2 - Feature #922: Add options -pS and -dl to application exiv2	Closed	25 Sep 2013

**Associated revisions**

---

**Revision 4166 - 05 Jan 2016 16:52 - Robin Mills**

#1108 and #1074 Correction to r4165 to fix MSVC build breaker and to document: exiv2 -eC (extract ICC profile).

**Revision 4168 - 07 Jan 2016 16:13 - Robin Mills**

#1108 and #1074 -pC and -pR (print Color Profile, print Recursively) now work on png.

**Revision 4171 - 07 Jan 2016 20:32 - Robin Mills**

#1108 and #1074 Correction to r4168 to fix MSVC build breaker.

**Revision 4224 - 11 Mar 2016 20:48 - Robin Mills**

#1108 Added photoshop/iptc parser to png/jpeg parser.

**Revision 4228 - 17 Mar 2016 22:14 - Robin Mills**

#1108 Added IPTC parser for tiff.

**Revision 4231 - 17 Mar 2016 23:45 - Robin Mills**

#1108. Refactored the IPTC printStructure code from png/jpeg/tiff into iptc.cpp

**Revision 4232 - 18 Mar 2016 07:48 - Robin Mills**

#1108 Refactored static indent(depth) from png/tiff/jpeg to Internal::indent(depth)

**Revision 4239 - 22 Mar 2016 16:17 - Robin Mills**

#1108 Corrections to test suite.

**Revision 4241 - 23 Mar 2016 13:49 - Robin Mills**

#1074 #1108 Added ICC profile to test/data/Reagan.jpg

**Revision 4285 - 22 Apr 2016 14:11 - Robin Mills**

#1108. Fixed issue with printing short strings which are stored in the directory offset field.

**Revision 4286 - 22 Apr 2016 14:30 - Robin Mills**

#1108. Correction to r4285

**Revision 4287 - 22 Apr 2016 15:13 - Robin Mills**

#1108. Correction to r4285. Code simplification.

**Revision 4295 - 02 May 2016 20:51 - Robin Mills**

#1108 Discovered another embedded tiff tag SubIFDs

**Revision 4497 - 15 Sep 2016 13:15 - Robin Mills**

#1108 Enhanced pngimage::printStructure() to display checksum

**Revision 4498 - 15 Sep 2016 14:09 - Robin Mills**

#1108 Better string formatting (and associated test/data changes). Tweaks to code layout for r4497.

**Revision 4503 - 16 Sep 2016 11:19 - Robin Mills**

#1108 Fixing issue with pngimage::printStructure() and the "Software" string in test/data/imagemagick.png

**Revision 4612 - 06 Oct 2016 19:52 - Robin Mills**

#1108 Added code to dump Exif, IPTC and iTXt/zTXt comment/description blocks for PNG files.

**Revision 4679 - 12 Nov 2016 22:05 - Robin Mills**

#1108 Documentation Update.

**Revision 4680 - 12 Nov 2016 22:08 - Robin Mills**

#1108 Fixed bugs in printStructure(kpsRecursive) handling of RATIONAL data.

**Revision 4682 - 13 Nov 2016 09:33 - Robin Mills**

#1108 Add test file for use in this document: [http://dev.exiv2.org/projects/exiv2/wiki/The\\_Metadata\\_in\\_TIFF\\_files](http://dev.exiv2.org/projects/exiv2/wiki/The_Metadata_in_TIFF_files)

**Revision 4683 - 13 Nov 2016 09:33 - Robin Mills**

#1108 Documentation update.

**Revision 4691 - 04 Dec 2016 06:02 - Robin Mills**

#1108 Add support to dump MakerNote IFDs with exiv2 -pR

**Revision 4694 - 04 Dec 2016 12:18 - Robin Mills**

#1108 exiv2 -pR to dump type == tiffIfd

**History**

---

**#1 - 27 Aug 2015 11:49 - Andreas Huggel**

Robin,  
  
You're looking at the small thumbnail image in IFD1 (Exif.Thumbnail.JPEGInterchangeFormat\*). That one is part of the Exif specs and we can deal with it just fine. The preview with the issue is one for which Sony add a tag to their makernote (Exif.Sony1.PreviewImage), which points to somewhere at the end of the file. It is a much larger preview (668kB in DSC01825.jpg), which doesn't fit into the Exif APP segment. Exiv2 first reads just the Exif APP segment and it cannot deal with anything outside of that segment easily later, so we currently just regard the tag value as invalid and truncate it. Worse, a subsequent Exif write operation e.g., adding an Exif tag to such an image will write the empty Exif.Sony1.PreviewImage tag back, so nobody will find the preview anymore afterwards.

**#2 - 27 Aug 2015 15:25 - Robin Mills**

- % Done changed from 0 to 10
- Estimated time set to 10.00 h

Thanks for this insight, Andreas.

The purpose of this issue is to dump more debug information about the file. And if there's some kind of orphan preview in the file, there probably nothing we can do. However recursively dumping the file (as above using dd) seems like a useful feature that can be implemented quite easily.

**#3 - 05 Jan 2016 15:08 - Robin Mills**

- % Done changed from 10 to 60

Andreas: You are right about Exif.Sony1.PreviewImage. That is a rather difficult subject. I re-encountered this over *The Holidays* when working on [#1143](#).

However the recursive dump that I am discussing here, is to extend the -pS feature (print Structure) to recursively descent all tiff-encoded IFDs which occur in JPG/APP1 Exif data, tiff files, following the tag ExifTag, the APP2/MPF data segment, some MakerNotes and (very likely) some other places that I have not yet discovered.

Here's the output of -pS (print Structure)

```
$ exiv2 -pS http://clanmills.com/Stonehenge.jpg
STRUCTURE OF JPEG FILE: http://clanmills.com/Stonehenge.jpg
address | marker | length | data
-----|-----|-----|-----
      2 | 0xd8 SOI |      0 |
      4 | 0xe1 APP1 | 15288 | Exif..II*.....
 15294 | 0xe1 APP1 |  2610 | http://ns.adobe.com/xap/1.0/.<?x
 17906 | 0xed APP13 |    96 | Photoshop 3.0.8BIM.....'.....
 18004 | 0xe2 APP2 |  4094 | MPF.II*.....0100.....
 22100 | 0xdb DQT |   132 |
 22234 | 0xc0 SOF0 |    17 |
 22253 | 0xc4 DHT |   418 |
 22673 | 0xda SOS |    12 |
$
```

And -pR (print Recursively):

```
$ exiv2 -pR http://clanmills.com/Stonehenge.jpg
STRUCTURE OF JPEG FILE: http://clanmills.com/Stonehenge.jpg
```

```

address | marker | length | data
2 | 0xd8 SOI | 0
4 | 0xe1 APP1 | 15288 | Exif..II*.....

```

STRUCTURE OF TIFF FILE (II): MemIo

address	tag	type	count	offset	value
10	0x010f Make	ASCII	18	146	NIKON CORPORATION
22	0x0110 Model	ASCII	12	164	NIKON D5300
34	0x0112 Orientation	SHORT	1	1	1
46	0x011a XResolution	RATIONAL	1	176	176/0
58	0x011b YResolution	RATIONAL	1	184	184/0
70	0x0128 ResolutionUnit	SHORT	1	2	2
82	0x0131 Software	ASCII	10	192	Ver.1.00
94	0x0132 DateTime	ASCII	20	202	2015:07:16 20:25:28
106	0x0213 YCbCrPositioning	SHORT	1	1	1
118	0x8769 ExifTag	LONG	1	222	222

STRUCTURE OF TIFF FILE (II): MemIo

address	tag	type	count	offset	value
224	0x829a ExposureTime	RATIONAL	1	732	732/0
236	0x829d FNumber	RATIONAL	1	740	740/0
248	0x8822 ExposureProgram	SHORT	1	0	0
260	0x8827 ISOSpeedRatings	SHORT	1	200	200
272	0x8830 SensitivityType	SHORT	1	2	2
284	0x9000 ExifVersion	UNDEFINED	4	808661552	
296	0x9003 DateTimeOriginal	ASCII	20	748	2015:07:16 15:38:54
308	0x9004 DateTimeDigitized	ASCII	20	768	2015:07:16 15:38:54
320	0x9101 ComponentsConfiguration	UNDEFINED	4	197121	
332	0x9102 CompressedBitsPerPixel	RATIONAL	1	788	788/0
344	0x9204 ExposureBiasValue	RATIONAL	1	796	796/0
356	0x9205 MaxApertureValue	RATIONAL	1	804	804/0
368	0x9207 MeteringMode	SHORT	1	5	5
380	0x9208 LightSource	SHORT	1	0	0
392	0x9209 Flash	SHORT	1	16	16
404	0x920a FocalLength	RATIONAL	1	812	812/0
416	0x927c MakerNote	UNDEFINED	3152	914	...

STRUCTURE OF TIFF FILE (II): MemIo

address	tag	type	count	offset	value
428	0x9286 UserComment	UNDEFINED	44	820	...
440	0x9290 SubSecTime	ASCII	3	12336	0.'
452	0x9291 SubSecTimeOriginal	ASCII	3	12336	0.'
464	0x9292 SubSecTimeDigitized	ASCII	3	12336	0.'
476	0xa000 FlashpixVersion	UNDEFINED	4	808464688	
488	0xa001 ColorSpace	SHORT	1	1	1
500	0xa002 PixelXDimension	SHORT	1	6000	6000
512	0xa003 PixelYDimension	SHORT	1	4000	4000
524	0xa005 InteroperabilityTag	LONG	1	4306	4306
536	0xa217 SensingMethod	SHORT	1	2	2
548	0xa300 FileSource	UNDEFINED	1	3	
560	0xa301 SceneType	UNDEFINED	1	1	
572	0xa302 CFAPattern	UNDEFINED	8	864	...
584	0xa401 CustomRendered	SHORT	1	0	0
596	0xa402 ExposureMode	SHORT	1	0	0
608	0xa403 WhiteBalance	SHORT	1	0	0
620	0xa404 DigitalZoomRatio	RATIONAL	1	872	872/0
632	0xa405 FocalLengthIn35mmFilm	SHORT	1	66	66
644	0xa406 SceneCaptureType	SHORT	1	0	0
656	0xa407 GainControl	SHORT	1	0	0
668	0xa408 Contrast	SHORT	1	0	0
680	0xa409 Saturation	SHORT	1	0	0
692	0xa40a Sharpness	SHORT	1	0	0
704	0xa40c SubjectDistanceRange	SHORT	1	0	0
716	0xa420 ImageUniqueID	ASCII	33	880	090caaf2c085f3e102513b2475

0041aa ...

130	0x8825 GPSTag	LONG	1	4060	4060
4338	0x0103 Compression	SHORT	1	6	6
4350	0x011a XResolution	RATIONAL	1	4426	4426/0
4362	0x011b YResolution	RATIONAL	1	4434	4434/0
4374	0x0128 ResolutionUnit	SHORT	1	2	2
4386	0x0201 JPEGInterchangeFormat	LONG	1	4442	4442
4398	0x0202 JPEGInterchangeFormatLeng	LONG	1	10837	10837
4410	0x0213 YCbCrPositioning	SHORT	1	1	1

15294 | 0xe1 APP1 | 2610 | http://ns.adobe.com/xap/1.0/.<?x

17906 | 0xed APP13 | 96 | Photoshop 3.0.8BIM.....'

18004 | 0xe2 APP2 | 4094 | MPF.II\*.....0100.....

STRUCTURE OF TIFF FILE (II): MemIo

address	tag	type	count	offset	value
---------	-----	------	-------	--------	-------

10		0xb000	MPFVersion		UNDEFINED		4		808464688		
22		0xb001	MPFNumberOfImages		LONG		1		3		3
34		0xb002	MPFImageList		UNDEFINED		48		52		...
22100		0xdb	DQT				132				
22234		0xc0	SOF0				17				
22253		0xc4	DHT				418				
22673		0xda	SOS				12				

\$

**#4 - 11 Mar 2016 20:49 - Robin Mills**

- % Done changed from 60 to 90

Added photoshop/iptc parser for png/jpeg files.

**#5 - 11 Mar 2016 21:13 - Robin Mills**

- Subject changed from *Recursively dump sub-files on an image.* to *Recursively dump sub-files of an image*

**#6 - 17 Mar 2016 22:15 - Robin Mills**

Added IPTC parser for tiff files.

**#7 - 28 Mar 2016 18:28 - Robin Mills**

- Status changed from *Assigned* to *Closed*

- % Done changed from 90 to 100

**#8 - 31 Mar 2016 13:56 - Robin Mills**

- Status changed from *Closed* to *Resolved*

- % Done changed from 100 to 80

**#9 - 22 Apr 2016 14:13 - Robin Mills**

- % Done changed from 80 to 90

[r4285](#). Fixed issue with printing short strings which are stored in the directory offset (dir[8:11]) field. This has been disturbing the output of the test harness for a while.

**#10 - 07 May 2016 13:08 - Robin Mills**

- Status changed from *Resolved* to *Closed*

- % Done changed from 90 to 100

- Estimated time changed from 10.00 h to 20.00 h

**#11 - 06 Oct 2016 19:53 - Robin Mills**

- Estimated time changed from 20.00 h to 26.00 h

[r4612](#) Added code to dump Exif, IPTC and iTXt/zTXt comment/description blocks for PNG files.

**#12 - 12 Nov 2016 22:09 - Robin Mills**

- Estimated time changed from 26.00 h to 32.00 h

[r4678r4679](#) Updated the TIFF documentation: [http://dev.exiv2.org/projects/exiv2/wiki/The\\_Metadata\\_in\\_TIFF\\_files](http://dev.exiv2.org/projects/exiv2/wiki/The_Metadata_in_TIFF_files)

[r4680](#) Fixed errors in printStructure(kpsRecursive) handling of RATIONAL.

**#13 - 04 Dec 2016 06:03 - Robin Mills**

- Estimated time changed from 32.00 h to 35.00 h