

Exiv2 - Feature #922

Add options -pS and -dl to application exiv2

25 Sep 2013 21:26 - Robin Mills

Status:	Closed	Start date:	25 Sep 2013
Priority:	Normal	Due date:	
Assignee:	Robin Mills	% Done:	100%
Category:	metadata	Estimated time:	15.00 hours
Target version:	0.26		
Description			
This is discussed in http://dev.exiv2.org/boards/3/topics/1608			
Related issues:			
Related to Exiv2 - Bug # 1081: Read XMP values from CR2 raw file when stored ...		Closed	15 May 2015
Related to Exiv2 - Feature # 1108: Recursively dump sub-files of an image		Closed	21 Aug 2015

Associated revisions

Revision 3650 - 27 Mar 2015 18:47 - Robin Mills

#922. Added options -pS and -pX to exiv2(.exe). Still to deal with -dl

Revision 3675 - 10 Apr 2015 18:50 - Thomas Beutlich

refs #922:

- Fix MSVC warning introduced by r3650
- Change first argument of Basiclo::seek to signed integer type

Revision 3702 - 17 Apr 2015 18:06 - Robin Mills

#922. Extract Extended XMP (multiple 65k block) and remove XMP blank lines.

Revision 3724 - 24 Apr 2015 13:28 - Robin Mills

#922. Work In Progress. Adding support for -pX and -pS for tiff files.

Revision 3726 - 24 Apr 2015 19:51 - Robin Mills

#922. Work in progress on options -pS and -pX

Revision 3728 - 25 Apr 2015 12:32 - Robin Mills

#922 -pS for TIFF tagName() uses Exiv2::exifTagList() (and similar) to find tag name.

Revision 3731 - 25 Apr 2015 20:17 - Thomas Beutlich

refs #922: Fix include and MSVC compilation

Revision 3737 - 26 Apr 2015 08:00 - Robin Mills

#1066. Fix for test exception. It's coming from #922 on all platforms except Mac.

Revision 3738 - 26 Apr 2015 09:50 - Robin Mills

#922. Documentation update. `Exiv2::Image::printStructure()` is not thread safe. No reason to use this in a multi-threaded application.

Revision 3744 - 27 Apr 2015 20:07 - Robin Mills

#922 -pS and -pX support for TIFF. Added formatters to Image class and use them from `{jpg/png/tiff}image.cpp`

Revision 3746 - 27 Apr 2015 20:34 - Robin Mills

#922. Don't remove blank lines from XMP. This is not Exiv2's business. -pX extracts XMP packet without modification.

Revision 3747 - 27 Apr 2015 21:25 - Robin Mills

#922. Fix Linux build breaker and MSVC compilation warnings.

Revision 3748 - 27 Apr 2015 21:33 - Robin Mills

#922. Fixing MSVC warnings.

Revision 3760 - 28 Apr 2015 19:19 - Robin Mills

#922. Fixing -pS and -pX on MSVC.

Revision 3768 - 30 Apr 2015 22:58 - Robin Mills

#922. Better platform and endian detection.

Revision 3769 - 01 May 2015 06:36 - Robin Mills

#922 Fixing `Image::formatString()` on Windows

Revision 3770 - 01 May 2015 06:54 - Robin Mills

#922. Mac fix for `Image::stringFormat()`

Revision 3771 - 01 May 2015 08:08 - Robin Mills

#922. Adding to the test suite.

Revision 3772 - 01 May 2015 09:16 - Robin Mills

#922 Rollback r3771. Very troublesome feature. bugfixed #922 is looping Linux. PNG Has endian issues on MM/PowerPC

Revision 3773 - 01 May 2015 11:36 - Robin Mills

#922 Submitting the fixed version of r3771

Revision 3781 - 02 May 2015 14:53 - Robin Mills

#1072 #922 BigEndian (Motorola PowerPC) fix.

Revision 4220 - 09 Mar 2016 07:51 - Robin Mills

#1057, #1064, #922, #1148. Work in progress. This is a composite patch of several matters in development. None are totally complete at this time.

Revision 4434 - 25 Aug 2016 16:20 - Robin Mills

#922 exiv2 -dl deletes all IPTC chunks in a JPEG.

Revision 4435 - 25 Aug 2016 18:32 - Robin Mills

#922 Correction to r4344 to handle msvc build breaker.

Revision 4436 - 25 Aug 2016 19:41 - Robin Mills

#922 Correction to r4434. Fixing another msvc build breaker.

History

#1 - 01 Sep 2014 12:34 - Robin Mills

- Assignee changed from Robin Mills to Tuan Nhu

I'm going to assign this to Tuan. It might make the v0.25 release.

#2 - 27 Mar 2015 18:48 - Robin Mills

- Category changed from iptc to metadata

- Assignee changed from Tuan Nhu to Robin Mills

Submitted r3650 to deal with -pS and -pX. Option -dl not yet implemented.

#3 - 30 Mar 2015 19:21 - Thomas Beutlich

After r3650 pngimage.cpp compiles with one warning on MSVC.

```
pngimage.cpp
..\src\pngimage.cpp(160): error C2220: warning treated as error - no 'object' file generated
..\src\pngimage.cpp(160): warning C4146: unary minus operator applied to unsigned type, result still unsigned
```

It is rather strange that the first argument of BasicIo::seek is an unsigned type in case of MSVC.

```
int seek(uint64_t offset, Position pos)
```

It either maps to `_fseeki64` or `std::fseek` which both take a signed type.

#4 - 30 Mar 2015 19:35 - Thomas Beutlich

- *File T922.patch added*

Attached patch should fix both issues.

#5 - 30 Mar 2015 20:21 - Robin Mills

Thanks for finding this Thomas. May I leave you to submit the patch when Andreas gets your SVN account set up. When you submit, you can take a look at Jenkins to see that it builds.

I'm rather doubtful about the FileIO object's ability to handle files > 3GB. It could even struggle with files > 1.5GB. This didn't matter much before the video code arrived. It's one of the matters I'd like the video guys to investigate.

#6 - 30 Mar 2015 20:21 - Robin Mills

- *Assignee changed from Robin Mills to Thomas Beutlich*

#7 - 30 Mar 2015 20:40 - Robin Mills

This is an interesting find, Thomas. I thought I had the option set in MSVC to say "treat warnings as errors". Perhaps I relaxed that condition during the integration of webready and did not restore it. I much prefer treating warnings as errors. Better to deal with warnings as soon as they surface.

I don't use that "treat warnings as errors" when I build the supporting libraries (expat, zlib, curl, ssh and openssl). I don't want to modify one line of the library code - so I ignore library build warnings.

You're welcome to raise an issue report about this and we can deal with it when you're done with the MemIO stuff, or when I return from vacation.

#8 - 30 Mar 2015 21:24 - Thomas Beutlich

It indeed should have never worked out for `blen != 0`. Do you have a test file and `exiv2` command line I can check with?

#9 - 10 Apr 2015 18:52 - Thomas Beutlich

- Assignee changed from Thomas Beutlich to Robin Mills

Patch submitted by r3675. Back to Robin for option -dl.

#10 - 17 Apr 2015 12:02 - Robin Mills

-pX is not implemented correctly when the XMP packet spans multiple segments. 652 rmills@rmillsmbp:~/gnu/exiv2/trunk/test \$ exiv2 -pX

data/exiv2-bug922.jpg

```

<x:xmpmeta xmlns:x="adobe:ns:meta/" x:xmptk="Adobe XMP Core 5.1.0-jc003">
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
    <rdf:Description rdf:about=""
      xmlns:GFocus="http://ns.google.com/photos/1.0/focus/"
      xmlns:GImage="http://ns.google.com/photos/1.0/image/"
      xmlns:GDepth="http://ns.google.com/photos/1.0/depthmap/"
      xmlns:xmpNote="http://ns.adobe.com/xmp/note/"
      GFocus:BlurAtInfinity="0.018819768"
      GFocus:FocalDistance="16.068678"
      GFocus:FocalPointX="0.4351852"
      GFocus:FocalPointY="0.39444447"
      GImage:Mime="image/jpeg"
      GDepth:Format="RangleInverse"
      GDepth:Near="10.917767524719238"
      GDepth:Far="38.58317565917969"
      GDepth:Mime="image/png"
      xmpNote:HasExtendedXMP="B9AC266F143C5DB7510D1CBEC51C924A"/>
  </rdf:RDF>
</x:xmpmeta>

```

B9AC266F143C5DB7510D1CBEC51C924A

... deleted ...

B9AC266F143C5DB7510D1CBEC51C924A

653 rmills@rmillsmbp:~/gnu/exiv2/trunk/test \$

xmpNote:HasExtendedXMP is documented on page 20 of this document:

<http://www.images.adobe.com/content/dam/Adobe/en/devnet/xmp/pdfs/XMPSpecificationPart3.pdf>

In researching this, I've discovered that XMP packets can be blank padded to enable applications to modify the XMP in-line. What a good idea! *Adobe people are really smart.* I'll probably update -pX output to remove trailing blanks.

#11 - 17 Apr 2015 18:10 - Robin Mills

#3702. Extract Extended XMP (spans more than one 65k segment). Remove blank lines. I still have to implement -dl.

#12 - 27 Apr 2015 20:31 - Robin Mills

r3744 removed the stripping of blank lines in the XML. It's better not to do this in Exiv2. An external tool such as xmllint can deal with this. -pX

extracts the XMP packet "as is" with blank lines.

#13 - 09 May 2015 08:03 - Robin Mills

- Target version changed from 0.25 to 0.26

-pS has been implemented for v0.25

-dl will not be implemented in v0.25 because I'm overloaded.

Here's the discussion with Jerome concerning -dl

<http://dev.exiv2.org/boards/3/topics/1608?r=1624#message-1624>

#14 - 23 May 2015 08:42 - Robin Mills

- Assignee deleted (Robin Mills)

#15 - 12 Oct 2015 09:03 - Robin Mills

- Assignee set to Robin Mills

#16 - 04 Jan 2016 09:53 - Robin Mills

- % Done changed from 0 to 50

- Estimated time set to 15.00

#17 - 21 Aug 2016 14:27 - Ben Touchette

I have a partially working implementation for -dl, but now taking a break for the rest of the day :)

#18 - 21 Aug 2016 15:22 - Robin Mills

- File ETH0138028.jpg added

- Assignee changed from Robin Mills to Ben Touchette

That would be wonderful. -pS -pR are done. The code for -dl is mostly done. Our old friend Stonehenge.jpg has a single IPTC block. 872

```
rmills@rmillsmbp:~/gnu/exiv2/trunk $ exiv2 -pS ~/Stonehenge.jpg
```

```
STRUCTURE OF JPEG FILE: /Users/rmills/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	0xe2 APP2	40940...4.....
22004	0xed APP13	96	Photoshop 3.0.8BIM.....!.....
22102	0xe2 APP2	4094	MPF.II*.....0100.....
26198	0xdb DQT	132	
26332	0xc0 SOF0	17	
26351	0xc4 DHT	418	
26771	0xda SOS	12	

```
873 rmills@rmillsmbp:~/gnu/exiv2/trunk $
```

-dl finds and reports the block. 871 rmills@rmillsmbp:~/gnu/exiv2/trunk \$ exiv2 -dl ~/Stonehenge.jpg

```
iptc data blocks: FOUND
```

```
17906 96
```

```
872 rmills@rmillsmbp:~/gnu/exiv2/trunk $
```

No JPEG should have more than one IPTC block - however a user had such a thing. I agreed to add -dl to clean it up for him. The code for -dl is in in

```
r4220 when it reports the blocks to delete in src/jpgimage.cpp      4220 robinwmill      if ( option == kpslptcErase ) {
    4220 robinwmill          std::cout << "iptc data blocks: " << (iptcDataSegs.size() ? "FOUND" : "none") << std::endl;
    4220 robinwmill          uint32_t toggle = 0 ;
    4220 robinwmill          for ( Uint32Vector_i it = iptcDataSegs.begin(); it != iptcDataSegs.end() ; it++ ) {
    4220 robinwmill              std::cout << *it ;
    4220 robinwmill              if ( toggle++ % 2 ) std::cout << std::endl; else std::cout << ' ' ;
    4220 robinwmill          }
    4220 robinwmill      }
```

Now that you know what basiclo does, you're going to have to rewrite the file without the "dead" blocks. Here's the confession in the log: 870
rmills@rmillsmbp:~/gnu/exiv2/trunk \$ svn log --revision 4220

```
-----
r4220 | robinwmills | 2016-03-09 07:51:04 +0000 (Wed, 09 Mar 2016) | 1 line
```

#1057, #1064, #922, #1148. Work in progress. This is a composite patch of several matters in development. None are totally complete at this time.

```
-----
871 rmills@rmillsmbp:~/gnu/exiv2/trunk $
```

I attach the bandit file with two IPTC blocks: 875 rmills@rmillsmbp:~/gnu/exiv2/trunk \$ exiv2 -pS ETH0138028.jpg

STRUCTURE OF JPEG FILE: <http://dev.exiv2.org/attachments/download/525/ETH0138028.jpg>

address	marker	length	data
2	0xd8 SOI	0	
4	0xe0 APP0	16	JFIF.....,.....
22	0xe1 APP1	5372	Exif..MM.*.....
5396	0xe1 APP1	7186	http://ns.adobe.com/xap/1.0/.<?x
12584	0xed APP13	18072	Photoshop 3.0.8BIM.....
30658	0xed APP13	18064	Photoshop 3.0.8BIM.....
48724	0xe2 APP2	576	ICC_PROFILE.....0ADBE....mnrRG
49302	0xee APP14	14	Adobe.d@.....
49318	0xdb DQT	132	
49452	0xc0 SOF0	17	
49471	0xdd DRI	4	
49477	0xc4 DHT	418	
49897	0xda SOS	12	

```
876 rmills@rmillsmbp:~/gnu/exiv2/trunk $
```

I'll assign this to you. If you get stuck, just assign it back to me.

If you're interested, you may wish to read the forum discussion to understand how the file with two IPTC blocks was created.

<http://dev.exiv2.org/boards/3/topics/1608?r=1624#message-1624>

#19 - 21 Aug 2016 15:33 - Ben Touchette

Thanks, already read the forum posts, right after reading this item from the todo list. Will do, right now i have added a set of function to toggle flag for the image to tell writeMetaData to disable writing app13_ that way i can call it from the Erase action, problem i have right now is i'm either overwriting or underwriting some data not sure and will require a bit more investigation and tinkering but those two APP13 blocks are gone. Also need to read more on the jpg format to see where i may have messed up. because now i end up with APP2 blocks for the problem file.

#20 - 21 Aug 2016 15:49 - Robin Mills

That's an interesting approach that I hadn't considered. I was just going to remove the blocks. I intended to binary copy the file leaving out the dead blocks. Can be done "in-line" in printStructure without bothering with readMetadata() or writeMetadata().

There's always more than one way to do things of course - however that was how I intended to deal with it.

#21 - 21 Aug 2016 15:56 - Ben Touchette

I thought about that as well when I first saw what you had already added in but then it dawned on me that this could be useful in the future for other formats in case something similar happens. It's a bit more complex but should provide more flex in future since there will already be a mechanism to forward that info to the write function. Copy pasting all those functions too each file format and renaming them took most of the morning.

#22 - 21 Aug 2016 16:07 - Robin Mills

Well, if you think that's the way to do this, please do that. If you discover that kpslptcErase is no longer used, could you totally remove it from the code base:

```
917 rmills@rmillsmbp:~/gnu/exiv2/trunk $ find . -name "*.?pp" -exec grep -H kpslptc {} \;
./include/exiv2/image.hpp:          , kpslccProfile , kpslptcErase
./src/actions.cpp:      rc = printStructure(std::cout,Exiv2::kpslptcErase);
./src/actions.hpp:      @brief Print image Structure information (used by ctlptcRaw/kpslptcErase)
./src/jpgimage.cpp:      if ( bPrint || option == kpsXMP || option == kpslccProfile || option == kpslptcErase ) {
./src/jpgimage.cpp:          // and dumping the XMP in a post read operation similar to kpslptcErase
./src/jpgimage.cpp:      } else if ( option == kpslptcErase && std::strcmp(signature,"Photoshop 3.0") == 0 ) {
./src/jpgimage.cpp:      if ( option == kpslptcErase ) {
./src/webpimage.cpp:      if ( bPrint || option == kpsXMP || option == kpslccProfile || option == kpslptcErase ) {
889 rmills@rmillsmbp:~/gnu/exiv2/trunk $
```

I was thinking we might want more delete/print functions to printStructure() in future. However let's stick exactly to the spec of this issue and implement -dl.

Incidentally, there's something similar to be fixed because -iX is removing the MakerNote #1064.

Thank You for getting involved. I'm so exhausted with Exiv2. You are getting the wind back in my sails. Thank You.

#23 - 21 Aug 2016 17:50 - Robin Mills

I've thought of something else about this. I believe it's a spec violation to have two IPTC blocks in a JPEG (and probably other formats that support IPTC). Simply saying "don't write IPTC", or "don't write JPEG/APP3/PhotoShop" might not be the correct fix. I have in mind that -dl will **exterminate** IPTC totally from the file. The option -dX exterminates XMP blocks.

And let me explain my thinking which are not **tablets of stone**. You are welcome to think differently about this.

The origin of printStructure() was work done by Tuan on the webready project. Tuan was a GSoc student. Very clever and able young man. Alison are going on vacation with him in Vietnam in December. He was curious about how the metadata was stored in the files. He added JpegImage::printStructure() and PngImage::printStructure() which he accessed from option --struct in exifprint.cpp. This was so useful, it was made available as **exiv2 -pS** in v0.25 which also had -pX and TiffImage::printStructure(). For v0.26, I added -pR and support dumping for IPTC blocks. Maybe v0.27 -pR will dump MakerNotes.

Exiv2 has supported for a long time -{i|d|e}tgt where **tgt: {a|e|i|x}+**. These options operate on the file foo.exv (which is a pure metadata jpg).

I've been thinking to support -{p|d}TGT where **TGT: {E|i|X|C}-}+**.

This is why -ix and -iX are different. **exiv2 -ix foo.abc** reads XMP from foo.exv. **exiv2 -iX foo.abc** reads XML from foo.xml

TGT: **-i-** means "read from stdin". TGT: **-e-** means "write to stdout" instead of foo.exv.

I hope you find that this explanation helpful. The essential point I want to make is that I had in mind that **-dl** would be a file operation that doesn't involve readMetadata() or writeMetadata() and is a file maintenance/repair feature. If I have confused you, just proceed as you think best!

#24 - 21 Aug 2016 18:05 - Ben Touchette

I'm not exactly my married to my idea lol. Definitely gives me stuff to ponder.

#25 - 21 Aug 2016 18:19 - Ben Touchette

Glad i could help it's tough to work on a project mostly solo; i know, i also think i need to reread the jpg specs. I'll be taking a closer look at it in the morning. Vietnam eh, sounds interesting i take it that it's a first time visit there?

#26 - 21 Aug 2016 18:51 - Robin Mills

It's good to work alone. I get to do what I think is best with little interference. On the other hand, when I'm swamped with too much to do, I get discouraged. Gilles insisting that we do WebP for v0.26 is a good example of interference. The priority is to finish v0.26, not to start new stuff. Anyway, WebP is done and now you're helping me to finish v0.26. So everything's going well.

Item 1 on our bucket list was to do huge things to our house. We're almost done. Item 2 is a "round the world" trip. Tuan comes from Vietnam and lives in Singapore. We'll visit Team Exiv2 members along the way. Christmas and New Year with Australian friends in Melbourne. We'll make our first visit to India, Vietnam, Singapore, NZ and Peru. <http://clanmills.com/BucketList.shtml>

#27 - 21 Aug 2016 19:01 - Ben Touchette

Indeed, and nice set of places to visit. I had Australia and NZ on my list, maybe thailand as well for S-E Asia.

#28 - 25 Aug 2016 16:19 - Robin Mills

- Assignee changed from Ben Touchette to Robin Mills
- % Done changed from 50 to 100

Ben isn't feeling very well, so I've taken this over again. Fix submitted r4434. I don't want to put the 1.69mb JPEG into the test suite. It is attached to this issue.

Ben if you ever want to undo my fix and apply your own, I'll be happy to review and discuss your version of fixing this. \$ cp ETH0138028.jpg

E.jpg ; exiv2 -pS E.jpg ; exiv2 -dl E.jpg ; exiv2 -pS E.jpg

STRUCTURE OF JPEG FILE: E.jpg

address	marker	length	data
2	0xd8 SOI	0	
4	0xe0 APP0	16	JFIF.....,.....
22	0xe1 APP1	5372	Exif..MM.*.....
5396	0xe1 APP1	7186	http://ns.adobe.com/xap/1.0/.<?x
12584	0xe2 APP2	576	rRGB XYZacspAPPL..
13164	0xed APP13	18072	Photoshop 3.0.8BIM.....
31238	0xed APP13	18064	Photoshop 3.0.8BIM.....
49304	0xe2 APP2	576	ICC_PROFILE.....0ADBE....mnrRG

```
49882 | 0xee APP14 | 14 | Adobe.d@.....
49898 | 0xdb DQT | 132
50032 | 0xc0 SOF0 | 17
50051 | 0xdd DRI | 4
50057 | 0xc4 DHT | 418
50477 | 0xda SOS | 12
```

Warning: JPEG format error, rc = 5

STRUCTURE OF JPEG FILE: E.jpg

```
address | marker | length | data
2 | 0xd8 SOI | 0
4 | 0xe0 APP0 | 16 | JFIF.....,.....
22 | 0xe1 APP1 | 5372 | Exif..MM.*.....
5396 | 0xe1 APP1 | 7186 | http://ns.adobe.com/xap/1.0/.<?x
12584 | 0xe2 APP2 | 576 | ....none.....
13164 | 0xe2 APP2 | 576 | rRGB XYZ .....acspAPPL..
13742 | 0xed APP13 | 576 | ICC_PROFILE.....0ADBE....mnrRG
14320 | 0xee APP14 | 14 | Adobe.d@.....
14336 | 0xdb DQT | 132
14470 | 0xc0 SOF0 | 17
14489 | 0xdd DRI | 4
14495 | 0xc4 DHT | 418
14915 | 0xda SOS | 12
```

\$ cp ~/Stonehenge.jpg S.jpg ; exiv2 -pS S.jpg ; exiv2 -dl S.jpg ; exiv2 -pS S.jpg

STRUCTURE OF JPEG FILE: S.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 | 0xe2 APP2 | 4094 | .....0...4.....
22004 | 0xed APP13 | 96 | Photoshop 3.0.8BIM.....'.....
22102 | 0xe2 APP2 | 4094 | MPF.II*.....0100.....
26198 | 0xdb DQT | 132
26332 | 0xc0 SOF0 | 17
26351 | 0xc4 DHT | 418
26771 | 0xda SOS | 12
```

STRUCTURE OF JPEG FILE: S.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 | 0xe2 APP2 | 4094 | .....0...4.....
22004 | 0xe2 APP2 | 4094 | .....0...4.....
26100 | 0xed APP13 | 4094 | MPF.II*.....0100.....
30196 | 0xdb DQT | 132
30330 | 0xc0 SOF0 | 17
30349 | 0xc4 DHT | 418
30769 | 0xda SOS | 12
```

614 rmills@rmillsmbp:~/gnu/exiv2/trunk \$

#29 - 27 Aug 2016 21:47 - Robin Mills

- Status changed from Assigned to Closed

Files

T922.patch	2.81 KB	30 Mar 2015	Thomas Beutlich
ETH0138028.jpg	1.69 MB	21 Aug 2016	Robin Mills