

Exiv2 - Feature #689

Support for Encapsulated PostScript (*.eps) files

24 Mar 2010 16:40 - Volker Grabsch

Status:	Closed	Start date:	24 Mar 2010
Priority:	Normal	Due date:	
Assignee:	Volker Grabsch	% Done:	100%
Category:	image format	Estimated time:	0.00 hour
Target version:	0.22		

Description

In our work, we (Michael Ulbrich and me) notice an increasing demand for handling metadata of Encapsulated PostScript files. While we could try to use Exiftool for that purpose, we would like to go another route and extend Exiv2 for that purpose.

Are you interested in such a contribution? Would you include it in your next Exiv2 release?

Has there been any unpublished work on that issue, which we could use as a base for our work?

Associated revisions

Revision 2479 - 17 Jun 2011 06:52 - Volker Grabsch

#689: provide support for EPS files

EPS is added to the front of the registry to ensure that very small EPS files will be detected.

This implementation refuses to deal with "read-only XMP" as well as "%Include..." DSC comments, because it is unclear how to handle those properly. If one of these special cases occur, a warning will be emitted which encourages the user to provide a real-world EPS file to us.

Revision 2482 - 18 Jun 2011 09:12 - Volker Grabsch

#689: Fixed regression: Ensure that `isEpsType()` doesn't disturb other file formats, and that none of the other file formats disturbs `isEpsType()`

Revision 2484 - 18 Jun 2011 16:34 - Volker Grabsch

#689: Provide support for more EPS variants

Revision 2485 - 18 Jun 2011 16:35 - Volker Grabsch

#689: Improved debug output in `EpsImage`

Revision 2486 - 19 Jun 2011 04:03 - Volker Grabsch

#689: Improved coding style and warnings in `EpsImage`

Revision 2487 - 19 Jun 2011 18:10 - Volker Grabsch

#689: Improved EPS tests

Revision 2488 - 19 Jun 2011 18:10 - Volker Grabsch

#689: Workaround for handling Exiftool's XMP embedding into EPS

Revision 2489 - 19 Jun 2011 18:10 - Volker Grabsch

#689: Improved performance of EpsImage

Revision 2491 - 19 Jun 2011 19:44 - Volker Grabsch

#689: Improved EPS tests

Revision 2492 - 19 Jun 2011 20:37 - Volker Grabsch

#689: provide an example EPS file that contains read-only XMP metadata

Revision 2493 - 19 Jun 2011 20:49 - Volker Grabsch

#689: provide an example EPS file that contains an %%Include DSC comment

Revision 2496 - 20 Jun 2011 07:15 - Volker Grabsch

#689: Improved EPS tests

Revision 2497 - 20 Jun 2011 07:33 - Volker Grabsch

#689: Added nested EPS files to the tests

Revision 2498 - 20 Jun 2011 08:30 - Volker Grabsch

#689: Bugfix for Photoshop EPS files

Revision 2499 - 22 Jun 2011 05:55 - Volker Grabsch

#689: Improved eps-test driver

Revision 2500 - 22 Jun 2011 05:56 - Volker Grabsch

#689: Improved Exiv2 EPS comments

Revision 2501 - 22 Jun 2011 05:56 - Volker Grabsch

#689: Workaround for Photoshop EPS files, adjusted test suite

Revision 2502 - 22 Jun 2011 05:56 - Volker Grabsch

#689: Improved code readability of EpsImage

Revision 2503 - 22 Jun 2011 06:32 - Volker Grabsch

#689: Add test EPS file that contains XMP metadata > 64 KB

Revision 2504 - 23 Jun 2011 04:12 - Volker Grabsch

#689: Improved block indentation

Revision 2505 - 23 Jun 2011 04:12 - Volker Grabsch

#689: Provide support for JPEG previews stored in the XMP metadata

Revision 2507 - 24 Jun 2011 07:59 - Volker Grabsch

#689: Support XMP previews that use the old xapGImg: namespace prefix

Revision 2508 - 24 Jun 2011 07:59 - Volker Grabsch

#689: Add EPS preview tests

Revision 2509 - 24 Jun 2011 10:33 - Volker Grabsch

#689: Make private implementation of EpsImage really private

Revision 2510 - 24 Jun 2011 15:03 - Volker Grabsch

#689: Bugfix in EPS test suite

Revision 2511 - 24 Jun 2011 15:03 - Volker Grabsch

#689: Added support for DOS EPS files

Revision 2514 - 27 Jun 2011 06:58 - Volker Grabsch

#689: Improved coding style of EpsImage: prefer byte* over char*, make use of getULong(), etc

Revision 2515 - 27 Jun 2011 07:03 - Volker Grabsch

#689: Improved coding style of EpsImage

Revision 2518 - 27 Jun 2011 09:39 - Volker Grabsch

#689: Bugfix in EPS test suite regarding file extensions for previews

Revision 2519 - 27 Jun 2011 09:39 - Volker Grabsch

#689: Added support for native DOS EPS previews

Revision 2585 - 24 Aug 2011 21:13 - Volker Grabsch

#689: Added support for nested EPS documents

History

#1 - 24 Mar 2010 19:48 - Andreas Huggel

Volker, Michael,

IIRC, support for EPS format has been requested before, so yes, there is interest and I'll be more than happy to include such a contribution provided it follows the usual conventions, but you've done that before, no worries there. I'm not aware of any existing work in this area or anybody else working on this. So all you have as a basis within Exiv2 is the code for the existing image formats.

Andreas

#2 - 02 Apr 2010 22:29 - Andreas Huggel

Volker, Michael, would it make things easier for you if you had write access to the repository?

Andreas

#3 - 02 Apr 2010 22:45 - Andreas Huggel

- Assignee changed from Andreas Huggel to Volker Grabsch

#4 - 04 Apr 2010 08:43 - Volker Grabsch

Andreas Huggel wrote:

| Volker, Michael, would it make things easier for you if you had write access to the repository?

Thanks for your trust. Having write access would indeed simplify things for us.

Note, however, that we can't start working on this right now. We'll come back to you.

(It seems that some applications such as "Adobe Indesign CS 2" don't preserve EPS metadata, so we'll have to investigate these issues first. It doesn't make sense to write metadata that will be stripped later in the workflow.)

Greetings,
Michael and Volker

#5 - 25 May 2010 20:13 - Andreas Huggel

- *Priority changed from High to Normal*
- *Target version deleted (0.20)*

#6 - 20 Mar 2011 15:24 - Volker Grabsch

I'd just like to announce that we finally started to work on this feature.

There is already a new class EpsImage which is able to read XMP metadata from certain kinds of *.eps images. I'll provide the patch as soon as we have support for writing and did some thorough testing.

This parser will probably also work for *.ps and *.ai files, because in the end of the day, these are all PostScript. However, we're currently concentrating on *.eps files as these are the most urgent need.

Also, especially *.ps files raise some questions regarding the design of the Exiv2 API. Exiv2 operates on image level, while *.ps files are documents. That is, they might have document-wide metadata, but also contain images which in turn might have metadata, or consist of other (sub) images. Strictly speaking, this issue also exists for *.eps files, but in EPS we are backed by the fact that the whole "document" is almost certainly meant to form a single image, and will usually consist of at most one single sub image.

So what do you think about the design issue regarding document/image formats?

Did this kind of issue already appear in other image formats or parsers?

#7 - 22 Mar 2011 18:12 - Andreas Huggel

Great! I've sent you and Michael SVN logins by mail so you can check-in your changes directly.

TIFF images can consist of a chain of IFDs with multiple images, called "multi-page TIFF". Usually these are related, eg. Exif uses this feature for the thumbnail and some RAW formats use it for copies of the same image with different resolutions. But it can also be used, eg., for different pages of a scanned fax, and each of the IFDs in the chain may have its own sub-IFDs.

Exiv2 struggles with these as the TIFF parser doesn't cater for all of this flexibility and its high-level design is for only one (set of) metadata container(s) per image/document as you said.

#8 - 15 Jun 2011 12:49 - Volker Grabsch

- *File exiv2-limited-eps-support-3.patch added*
- *Target version set to 0.22*
- *% Done changed from 0 to 80*

Attached is our current state of work. We'd like to commit it, but it would be great to get some feedback before doing so. As of now, the implementation has the following drawbacks:

- no support for "read-only XMP" (it seems that nobody uses this anyway)
- no support for "%%Include..." DSC comments (it seems that nobody uses those anyway)
- no support for deleting XMP metadata
- no support for DOS EPS (will be added later)
- no support for embedded documents (will be added later)
- no support for native previews that are stored as DSC comments (will be added later)
- no test cases (see below)

Is this state of work good enough to be committed?

Note that none of these issues will cause a crash. Each issue will be reported as warning or error in case it occurs.

Regarding the test cases: We have created more than 60 test files with various tools and various options. Since we created all these EPS files on our own, there shouldn't be any copyright issues with those.

However, we wonder how to integrate those into Exiv2. Should we simply add those to test/data/ ?

In addition, for good regression tests we probably need "before/after" variants for each file, but that would result in more than 120 files in total. Would that be okay for you?

Also, should we try to modify the existing test drivers such as "write-test.sh", or should we create our own test driver such as "eps.sh" for those tests?

Greetings,
Michael and Volker

#9 - 16 Jun 2011 08:47 - Andreas Huggel

| *Is this state of work good enough to be committed?*

Thanks for the patch! Please commit it. As soon as it's in the trunk, it will be exposed to some level of ongoing regression testing by all those who track the trunk, that's always good.

| *Note that none of these issues will cause a crash. Each issue will be reported as warning or error in case it occurs.*

Good.

| *Regarding the test cases: We have created more than 60 test files with various tools and various options. Since we created all these EPS files on our own, there shouldn't be any copyright issues with those.*

| *However, we wonder how to integrate those into Exiv2. Should we simply add those to test/data/ ?*

Or in a subfolder like test/data/eps? It's good to have a lot of test cases, but I have never dealt with so many in one go. Typically, I run these tests before, sometimes during and then again after any development. If something fails it should be as easy as possible to identify the test case and test data involved.

| *In addition, for good regression tests we probably need "before/after" variants for each file, but that would result in more than 120 files in total. Would that be okay for you?*

Ok, as long as they are reasonably small. Only a few KB each is best.

Also, should we try to modify the existing test drivers such as "write-test.sh", or should we create our own test driver such as "eps.sh" for those tests?

Please create new test drivers. I think that will be easier to maintain and troubleshoot. Generally, it's easier to troubleshoot problems if a test is done using the exiv2 utility as opposed to some C++ test-driver that runs lots of test cases.

Andreas

#10 - 17 Jun 2011 06:51 - Volker Grabsch

- File deleted (exiv2-limited-eps-support-3.patch)

#11 - 18 Jun 2011 09:19 - Volker Grabsch

Andreas Huggel wrote:

Thanks for the patch! Please commit it.

Done. (r2479 and r2482)

Or in a subfolder like test/data/eps? It's good to have a lot of test cases, but I have never dealt with so many in one go. Typically, I run these tests before, sometimes during and then again after any development. If something fails it should be as easy as possible to identify the test case and test data involved.

Okay, so the test cases went go into test/data/eps, the driver is test/eps-test.sh.

In addition, for good regression tests we probably need "before/after" variants for each file, but that would result in more than 120 files in total. Would that be okay for you?

Ok, as long as they are reasonably small. Only a few KB each is best.

I tried to make those files as small as possible, but some tools just don't export small EPS files. Those generate around 200 kb of EPS code even when exporting an empty document. I could try to shorten those by hand, but that might distort the tests.

For now, I committed only a portion of the tests. Those are about 3 MB (with additional 3 MB for expected outputs). Should I add some more test files? (around 4 MB more) What should be the maximum size for the total EPS test suite? (including expected outputs)

Please create new test drivers. I think that will be easier to maintain and troubleshoot.

Okay, so in r2479, I committed a first version of the test driver. More checks will be added soon.

Generally, it's easier to troubleshoot problems if a test is done using the exiv2 utility as opposed to some C++ test-driver that runs lots of test cases.

I was using the exiv2 utility for testing anyway.

Greets,
Volker

#12 - 19 Jun 2011 19:48 - Volker Grabsch

I just improved the tests as well as the EpsImage implementation. (r2484, r2485, r2486, r2487, r2488, r2489, and r2491)

Along the way, I fixed a bug in the exiv2 command line tool. (r2490)

Open issues are still DOS EPS, embedded documents, and native previews as well as XMP previews.

I'd like to add more test EPS files, but I need to know what you consider the maximum size of test/data/eps/.

#13 - 19 Jun 2011 23:35 - Andreas Huggel

Great. I suggest you go ahead and check in all test cases. Then we'll see if anybody complains. As the tests are not part of the distributed tarball, AFAICS the only occasion when they may be a problem is during the initial download.

#14 - 20 Jun 2011 15:18 - Volker Grabsch

Okay, so I improved the eps-test driver and added the remaining test files. (r2496, r2497)

I also added two test files that trigger the "read-only XMP" and the "%%Include..." issue, demonstrating the already mentioned special warning that is meant to encourage the user to send the file to us. (r2492, r2493)

Finally, I added one more fix for Photoshop EPS, but there are still issues with that. (r2498)

#15 - 23 Jun 2011 04:24 - Volker Grabsch

There is an additional big "WTF?!" in the way Photoshop handles its own EPS files, which required some workarounds. (r2501)

I also improved some minor issues. (r2499, r2500, r2502, r2503)

Finally, I added support for JPEG previews that are stored in the XMP metadata, which required a Base64 decoder. (r2504, r2505)

Open issues are: Native previews (i.e. those not stored in XMP), embedded documents, and DOS EPS.

To make licensing issues clear: I wrote that Base64 decoder on my own (r2505), inspired by existing free implementations that were all too complicated in my opinion. I implemented a simple algorithm that behaves well in all corner cases, so no extra code for strange "special cases" regarding padding etc. was needed. I took extra care to make the implementation safe against integer overflows, and of course against buffer overflows.

The only drawback is that it requires "int" to be at least 32-bit, so this code won't work on old 16-bit platforms. Is that an issue for Exiv2?

(Maybe I'll also release that decoder separately under BSD license or Public Domain ...)

#16 - 24 Jun 2011 15:11 - Volker Grabsch

After some minor improvements (r2507, r2508, r2509, r2510) I finally managed to add support for DOS EPS files. (r2511)

However, the contained TIFF or WMF preview is not yet exposed to the preview manager. Other open issues are: Native (DSC) previews, and nested documents.

#17 - 25 Jul 2011 08:53 - Volker Grabsch

- % Done changed from 80 to 90

Hello Andreas,

We added lots of other improvements to the EpsImage implementation, up to r2566.

Summary

The DOS EPS previews (TIFF, WMF) are exposed to the PreviewManager (r2519). We also decode Photoshop EPS previews, which are hex encoded IRBs as EPS comments (r2559, r2560). For that purpose, we make use of the Photostop struct as you [suggested in the Forum](#), extending it by a new method locatePreviewIrb (r2558). We added support for %%Include DSC comments (r2549), inspired by a real-world example that we found recently. We also added support for deleting XMP metadata from EPS files (r2547).

Remaining issues

- no support for nested EPS documents (*test EPS files are already present*)
- no support for DSC previews and Illustrator previews (*starting with %%BeginPreview and %AI7_Thumbnail, both are very similar*)
- no support for EPS files containing "read-only XMP" (*it is unclear how to handle those cleanly, and it seems that nobody uses those anyway*)

#18 - 17 Aug 2011 14:56 - Volker Grabsch

We just added support for Illustrator previews (%AI7_Thumbnail) in r2577.

However, we will not add support for DSC previews (%BeginPreview) as those are really flawed. The standard doesn't say anything about colors, so at most gray scale previews are possible via that DSC comment. Also, no Adobe application makes use of that. Only OpenOffice and Inkscape produce it, and those only when you explicitly enable an "EPSI" option. However, the OpenOffice EPSI produces only a 1-Bit black/white preview, so it is almost never useful (see test/data/eps/eps-nested_noxmp_oodraw-lev2-epsi.eps). Inkscape produces a more usable 8-Bit gray scale preview, but in EPSI mode it creates a very outdated kind of EPS that starts with "IPS-Adobe-2.0 EPSF-1.2" (see test/data/eps/eps-flat_inkscape-epsi.eps). To summarize, 1) the standard is flawed, and 2) no application produces an actually usable EPSI file that contains a useful DSC preview.

Also, we will not add support for "read-only XMP", because this is a purely theoretical option of the standard that isn't used by any real-world application. Even if it was used, it would not be clear how to handle that. However, in the very unlikely case that such an EPS file runs through Exiv2, we will issue a warning message that attempts to encourage the user to provide that EPS file to us. Until that happens, we shouldn't waste any more thought on this esoteric topic.

So the only remaining issue is:

- no support for nested EPS documents (*test EPS files are already present*)

#19 - 24 Aug 2011 21:16 - Volker Grabsch

- Status changed from New to Resolved

- % Done changed from 90 to 100

Superb and thank you very much! We have lots of public holidays here next week, I should have some time to release 0.22, that is overdue anyway. Or do you see any reason why we should hang on longer?

Andreas

#20 - 25 Aug 2011 03:54 - Volker Grabsch

A release would be great to get this code out to the people as soon as possible.

Maybe some compiling issues for MSVC need to be fixed, but apart from that I don't see any problems.

#21 - 29 Aug 2011 04:34 - Volker Grabsch

Somehow my comment about the last remaining issue vanished. So I'm writing it here again for clarity:

I finally added support for nested EPS documents. (r2585)

So this huge task is finished, there are no remaining issues left.

#22 - 18 Sep 2011 05:27 - Andreas Huggel

- Status changed from Resolved to Closed