Making patches; Attribution; Copyright and License.

Egon Willighagen

Published June 30, 2009

Citation

Willighagen, E. (n.d.). In *chem-bla-ics*. chem-bla-ics. https://doi.org/10.59350/dyxec-j7q96

Keywords

Opensource, Cdk

Copyright

Copyright © Egon Willighagen 2009. Distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

I have discussed this in the past on mailing lists, but realized yesterday that I need to strengthen the message a bit more. Just to remove confusion. The below is extracted from an email I sent this morning to the cdk-user mailing list, but I'm sure you can apply this to any other open source project. (Disclaimer: I have not studied international law, and the below cannot be used as legal advice. Like you would have! Hahahaha! Let it be pointers :)

1. What is that copyright/license header in that .java source file?\

This header looks something along the lines of:

```
/* Copyright (C) 2000-2007  Christoph Steinbeck
 *               2001-2007,2009  Egon Willighagen
 *
 * Contact: cdk-devel@lists.sourceforge.net
 *
 * This program is free software; you can redistribute it and/or
 * modify it under the terms of the GNU Lesser General Public License
 * as published by the Free Software Foundation; either version 2.1
 * of the License, or (at your option) any later version.
 * All we ask is that proper credit is given for our work, which includes
 * - but is not limited to - adding the above copyright notice to the beginning
 * of your source code files, and to any copyright notice that you may distribute
 * with programs based on this work.
 *
 * This program is distributed in the hope that it will be useful,
 * but WITHOUT ANY WARRANTY; without even the implied warranty of
 * MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
 * GNU Lesser General Public License for more details.
 *
 * You should have received a copy of the GNU Lesser General Public License
 * along with this program; if not, write to the Free Software
 * Foundation, Inc., 51 Franklin St, Fifth Floor, Boston, MA 02110-1301 USA.
 */
```

This header has two major pieces: 1. who has the copyright on this file; 2. what is the license that makes is Open Source.

This is crucial information, and the CDK has a bad history in keeping track of primarily 1. Many source files, actually, still list the Chemistry Development Kit Project has copyright owner, which is a false statement, as the CDK Project is not a legal entity and in many countries therefore not allowed to own copyright. Moreover, none of the contributors ever signed a legal paper to re-assign copyright to this project anyway, like we do with many of our ACS papers.

2. But doesn't the Git/SVN/CVS history have this copyright owner information?\

It is true that the Git, SVN and CVS histories of the CDK source code contain a lot of information on this. However, this is not helping, because this information is lost when we distribute our source code. And when others distribute our source code (e.g. Debian and Ubuntu), they have no means of keeping track of this.

Therefore, we must properly annotate our source files with this information.

3. Why is there a contact email?\

Because the CDK project should still be the single point of entry regarding the source code.

4. What's with all those years listed in those copyright ownership lines?\

Seriously, I have no clue, but every serious project does it, so there must be some legal reason. Indeed, it sounds logical to only list the years when you actually made changes. In the above hypothetical header, I made changes between 2001 and this year, but not in 2008. It might have to do with proper establishment of the end of the copyright period; see below.

5. When must I add my name to the copyright lines?\

When you made a non-trivial contribution to the source code, and you must ensure you do this in each such contribution. By adding your name, you make clear that:

a. you are the original author of that contribution (and not someone else)
b. you release the software under the given license

This is the information (re)distributors need to know if they are working within the boundaries of law.

6. Should I complain about people not adding this information in their patches when reviewing those contributions?\

Yes, you should.

7. What about all those files that still list Copyright (C) The CDK Project?\

File bug reports. For each file, we need to read the commit history, extract the authors of all non-trivial contributions and when those contributions were made, and update the copyright lines.

8. Must the header always list the LGPL as license?\

No. The LGPL is our license choice, but if you used code under another (compatible) license written by someone else, that original license applies, and that original license you need to provide in the header.

Additionally, do not forget to list the original copyright owners.

9. Can I rewrite GPL C++/C code as LGPL for the CDK?\

chem-bla-ics

Not entirely related to the above, but relevant. I once asked the FSF about this, and rewriting a piece of code in another language is *not* a clean room implementation and does, therefore, not erase original copyright ownership not license applicability. Therefore, we cannot base CDK implementations on, e.g. GPL-ed licensed C++ code, such as in OpenBabel.

10. How can I use GPL code in the CDK?\

You cannot. All code depending GPL-code, must be GPL too. There is the ChemoJava project to hold GPL-licensed CDK-based code, which has a number of classes that use the CDK library and depend on GPL libraries too.

Alternatively, you can try to convince the original authors to relicense. A good recent example here, is the UFF force field implementation of OpenBabel which was relicensed (or dual licensed) as LGPL and now also available from Jmol. Incidentally, this is a reason why it is important to have those copyright lines include the email address, so that in the future people can contact the original authors of code. In a far future, this is even used to decide when copyright no longer applies, because the original authors are dead by then.

(Thanx to Noel for replying to the mailing list!)