

LICENSING

“Good writers borrow from other writers. Great writers steal from them outright.

Aaron Sorkin



Why license?

- Software licenses are about two things: **defining copyright**, and **protecting yourself** from being held liable if your software screws something up somewhere
- In the US, copyright is automatic – the right to copy your code is yours, no one else can copy it

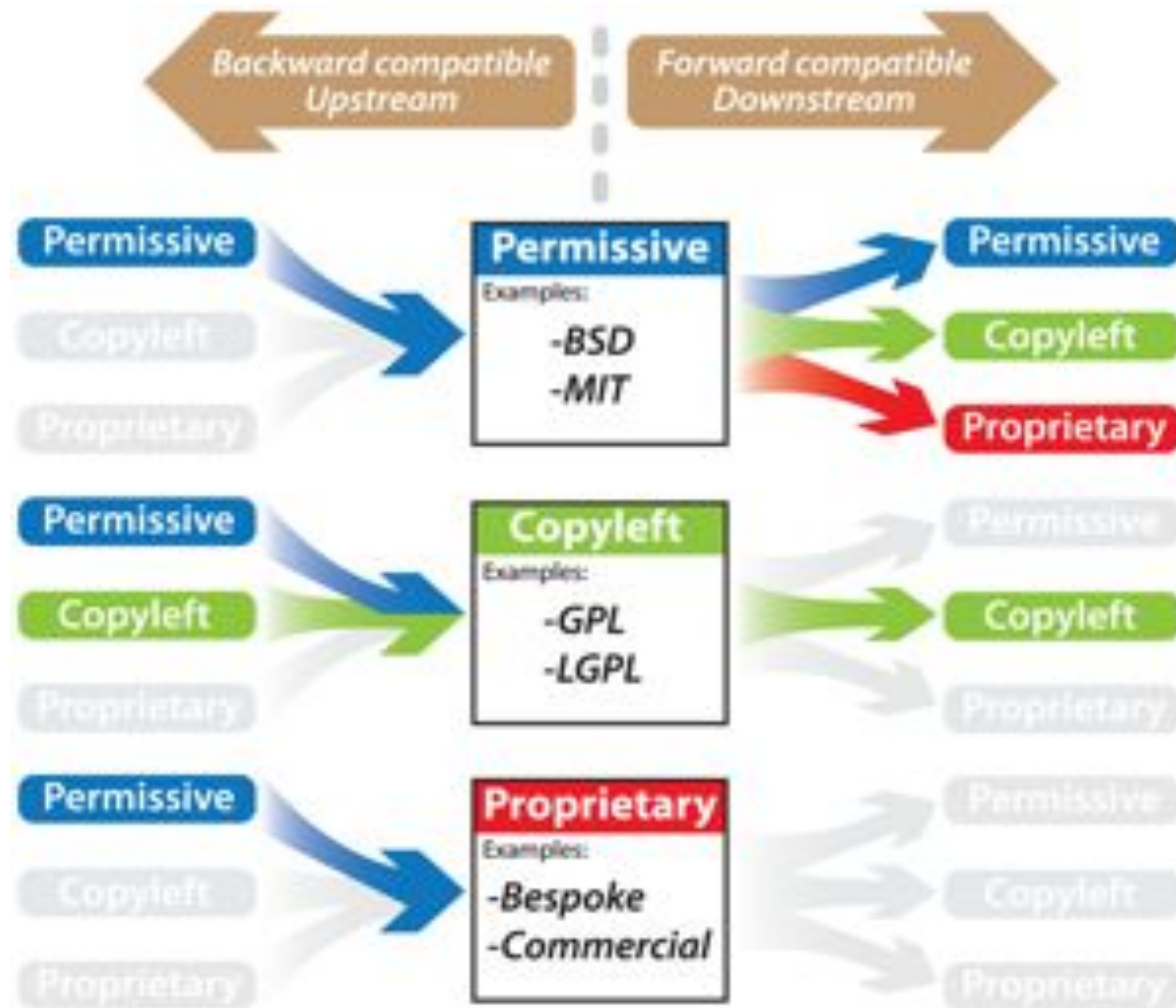
Copyright and Open Source

- Any creative work, including software code, is automatically eligible for intellectual property (and thus **copyright**) protection.
- Code that appears to be, or is expressly advertised as freely available has not waived such protection

The open licences certified by the Open Source Initiative (OSI) grant at least the following rights

- The source code is available, and may be used and redistributed without restrictions, including as part of aggregate distributions
- Modifications or other derived works are allowed, and can be redistributed as well
- The question of who receives these rights is not subject to discrimination, including not by fields of endeavor such as commercial versus academic

Figure 2. Schematic representation of license directionality.



Morin A, Urban J, Sliz P (2012) A Quick Guide to Software Licensing for the Scientist-Programmer. PLoS Comput Biol 8(7): e1002598. doi:10.1371/journal.pcbi.1002598
<http://journals.plos.org/ploscompbiol/article?id=info:doi/10.1371/journal.pcbi.1002598>

Software Licensing

A few licenses are by far the most popular, including the following:

- [GNU General Public License](#) (GPL),
- [MIT license](#),
- Do not use [Creative Commons](#) licenses for software ([CC FAQ](#))

A listing of more than 50 existing open-source licenses can be found on the Open Source Initiative (OSI) page (see <http://www.opensource.org/licenses/category>)

GPL-3

- Use, modify, distribute, ...
- Don't hold the author liable.
- Distributions must include the source code.
- Software incorporating the work must also be under GPL-3 (MIT does not require that)

Data Licensing

- In most jurisdictions most types of data (with possibly the exception of photos, medical images, etc) are considered facts of nature, and are hence not eligible for copyright protection
- Therefore, using a license is confusing and inappropriate
- It is much better to clarify the legal side with a public domain waiver such as [Creative Commons Zero \(CC0\)](#), and the social expectations side with express requests for how to use and cite the data

CC0 1.0 Universal (CC0 1.0) Public Domain Dedication

No Copyright



The person who associated a work with this deed has **dedicated** the work to the public domain by waiving all of his or her rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law.

You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission. See **Other Information** below.



Publications Licensing

Creative works: Manuals, reports, manuscripts and other creative works are eligible for intellectual property protection and are hence automatically protected by copyright, just as software source code

Creative Commons has prepared a set of licenses using combinations of four basic restrictions:

- **Attribution** (CC-BY): derived works must give the original author credit for their work
- **No Derivatives** (CC BY-ND): people may copy the work, but must pass it along unchanged
- **Share Alike** (CC-BY-SA): derivative works must license their work under the same terms as the original
- **Noncommercial** (CC BY-NC): free use is allowed, but commercial use is not
- **Mix-and-match** (CC BY-NC-SA, CC BY-NC-ND)

Only the Attribution (CC-BY) and Share-Alike (CC-BY-SA) licenses are considered “Open”

Use CC-BY for publications



Human subjects research

- If you do human subjects research, you can not just put the data out
- Human subjects research must be reviewed by an Institutional Review Board (IRB). Clear protocol, informed consent, data protection plan
- Anonymized data may be exempt. But the IRB makes determination
- Not everything is research, e.g., data used only in a course

HIPAA

- HIPAA = Health Insurance Portability and Accountability Act of 1996
- Special rules about medical data with any identifying information – focus on privacy and security
- Definition of “potentially identifying information” is very broad (zip code, dates of a survey)
- Special security measures = paperwork

How to add a license

When a repository with source code, a manuscript or other creative works becomes public, it should include a file **LICENSE** or **LICENSE.txt** in the base directory of the repository that clearly states under which license the content is being made available

You may also want to include a file called **CITATION** or **CITATION.txt** that describes how to reference your work

Example of CITATION

BibTex format

```
@online{wilson-software-carpentry-2013,  
  author      = {Greg Wilson},  
  title       = {Software Carpentry: Lessons Learned},  
  version     = {1},  
  date        = {2013-07-20},  
  eprinttype  = {arxiv},  
  eprint      = {1307.5448}  
}
```

Rendered as

Greg Wilson: "Software Carpentry: Lessons Learned". arXiv:1307.5448, July 2013.

Summary

- If you don't license your software, it can't be modified or reused - pick a license, any license
- Use MIT or GPL for software
- Use CC0 for data
- Cite sources of software and data
- Be careful with human data – ask, if not sure