

Why Copyleft Isn't Right for Scientific Code

Victoria Stodden

Assistant Professor

Department of Statistics

Columbia University

IP Scholars

UC Berkeley, California

August 12, 2010

Open Science is a Movement

Open Source Software

Reproducible Research Standard

Open Science is a Movement

Technology is driving the change to Open Science:

- ▶ computational pervasiveness across fields, and changing the nature of knowledge discovery,
- ▶ changing communication modalities,
- ▶ opportunities to conform more closely to the scientific method, ie. *reproducibility of computational results*.

Reproducibility requires open code and data \Rightarrow confronting copyright issues in code sharing.

The Importance of Code in Science

- ▶ Code is regularly produced in the process of scientific discovery and publication,
- ▶ Frequently embodies novel intellectual contributions,
- ▶ Often embodies myriad decisions necessary for replication of the results, for example data preparation,
- ▶ Required for the full understanding of the new idea,
- ▶ Open code more easily encourages building on previous ideas
- ▶ New journals dedicated to software cropping up..

Open Source Software Licensing Solutions

Open Source Software: software with licenses that communicate alternative terms of use to code developers, rather than the default assigned by copyright law.

Richard Stallman created the GNU Public License (GPL) in 1989 to ensure distribution of source code, with compiled programs. Majority of open source code under GPL.

Since then hundreds of software licenses have been created with varying terms:

- ▶ (Modified) BSD license
- ▶ MIT license
- ▶ Apache 2.0
- ▶ “Lesser” GPL v3
- ▶ ... (see <http://www.opensource.org/licenses/alphabetical>)

the GPL and Scientific Codes

Today we have:

- ▶ pervasiveness of the GPL in the Open Source Community spilling over into the scientific community,
- ▶ majority of posted scientific code released under GPL, if licensed at all.

Argument: Scientific codes exist in a different normative structure than open source software and require a licensing structure consonant with those norms.

Why Copyleft isn't Right for Scientific Code

- ▶ The scientific ethos precludes directing another scientist's creative contribution.
- ▶ Copyleft licenses make demands on downstream code, namely that they use the upstream license on the entire library of new code. Two codes under two different copyleft licenses therefore cannot be mixed, as code cannot carry two licenses.
- ▶ Science creates knowledge, our best estimate of the truth - a public good. People are free to use and build on public goods however they see fit.
- ▶ Society's right to knowledge is an unconditional right.
- ▶ Copyleft creates a barrier the transmission of knowledge and scientific progress, that is not compensated by other benefits.
- ▶ Copyleft inhibits or prevents collaboration with industrial partners.

Solution: Attribution-only Licensing for Scientific Code

Attribution-only licensing (or public domain) for scientific works:

- ▶ Reproducible Research Standard (Stodden, 2009),
 1. Release media components (text, figures) under CC BY,
 2. Release code components under Modified BSD or similar,
 3. Release data to public domain (CC0) or attribution license.
- ▶ Builds on the notion of *Research Compendium* (Gentleman & Lang, 2004).

References:

“Enabling Reproducible Research: Open Licensing for Scientific Innovation”

“15 Years of Reproducible Research in Computational Harmonic Analysis”

“The Legal Framework for Reproducible Research in the Sciences: Licensing and Copyright,”

“The Scientific Method in Practice: Reproducibility in the Computational Sciences”

<http://www.stanford.edu/~vcs>

Data and Code Sharing Roundtable, Nov 2009:

<http://www.stanford.edu/~vcs/RoundtableNov212009/>