

Legal Attribution and Academic Citation: Congruence in Licensing of Academic Research

Victoria Stodden

Fellow, Internet and Society Project

Yale Law School

Fellow, Science Commons

IP Scholars Conference

Cardozo School of Law

August 6, 2009

Agenda

Copyright and Academic Research - Quick Tour

Copyright Opposes Scientific Norms

Reproducible Computational Research

Principle of Academic Licensing

Licensing in Academia

Summary and Other Considerations

Copyright Opposes Scientific Norms

- ▶ Scientific norms imply openness in all aspects of research.
- ▶ Copyright falls by default on original expressions such as academic authorship, papers, code, and aspects of dataset creation.
- ▶ Thus, default IP structure for academic work operates counter to scientific norms.

The Reproducible Research Standard (Stodden 2009)

Goals:

- ▶ realign IP rights with scientific norms,
- ▶ release of all scientific research, including code and data.

To satisfy the Reproducible Research Standard:

1. CC BY on media such as text, figures,
2. Attribution license on code: such as Apache 2.0, MIT, LGPL,
3. Data under CC0 or Science Commons Open Access Data Protocol,
4. "original selection and arrangement" of the data, under CC BY or attribution open source license.

Principle of Academic Licensing

Principle of Scientific Licensing: Legal encumbrances to the dissemination, sharing, use, and re-use of scientific research compendia should be minimized, each requiring a strong and compelling rationale before application.

(Stodden, *Enabling Reproducible Research: Open Licensing for Scientific Innovation*, 2009, p. 14.)

Using a Creative Commons License in Academia

For CC BY licenses, attribution is a condition of use:

- ▶ Section 4. b. of CC BY 3.0:
- ▶ "You must . . . keep intact all copyright notices for the Work and provide, *reasonable to the medium* or means You are utilizing:" [emphasis added]
 - ▶ Author Name
 - ▶ Work Title
 - ▶ License URI, as practicable
 - ▶ Credit for any adaptation

In academic settings, credit is given through local norms-based citation practices. Is a reconciliation possible?

Using a Creative Commons License in Academia

When CC licensing doesn't add extra work:

- ▶ Academics must cite according to convention which includes Author Name, Title of Work.
- ▶ "reasonable to the medium" language suggests this could satisfy CC BY. (Think Nguyen)

When CC licensing does add extra work:

- ▶ Including the license notice is new.

Using an (Attribution) Open Source Software License in Academia

For Reproducible Research Standard recommended licenses, attribution is a condition of use:

- ▶ Copyright Statement must be included with the work (all licenses),
- ▶ List of contributors,
- ▶ Notice of license, copy of license(s) (LGPL).

Local norms-based citation practices may satisfy contributor lists, but other requirements of licenses are an additional burden to the scientist distributing code.

Could Tagging and Tools Minimize the Burden of OS License Compliance?

HTML tags could contain:

- ▶ license information,
- ▶ links to license text,
- ▶ author information,
- ▶ other descriptive meta-data for the work.

Tools exist to do this:

- ▶ automatic meta-data extraction from academic papers,
- ▶ searching HTML+RDFa tags (e.g. Yahoo!'s SearchMonkey)
- ▶ web interface tools: ie. CC license generation tool.

Impact

1. Copyright notices and long attribution streams automatically generated in tag format imply license compliance, including attribution-only open source code licenses.
2. New metric for measuring academic performance:
 - ▶ takes data and code contributions into account, as well as citations,
 - ▶ finds scientific contributions beyond Web of Science.
3. Semantic web markup implemented as by-product.
4. Potential increase in individual attribution.

Summary and Other Considerations

- ▶ Legal questions remain:
 1. How far can we push “reasonable to the medium?”
 2. Does an HTML tag suffices for license compliance: giving a link to the terms, providing a copyright notice, attribution.
 3. Copyright thickets through the promotion of individual rights assignment? Rights-clearing mechanism, reproducibility certification...
- ▶ Auxiliary tools aren't sufficient (yet) to supplant the additional burden due to license compliance.

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

Email: vcs@stanford.edu