

The Reproducible Research Standard: Copyright and the Scientific Method

Victoria Stodden

Yale ISP Tuesday Lunch Talk
February 10, 2009

Massive Computation

Emblems of our age include:

- data mining for subtle patterns in vast databases;
- massive simulations of a physical system's complete evolution repeated numerous times, as simulation parameters vary systematically.

Computing and Science

- Fundamental transformation in the way we do science.
- Computation emerging as a third branch of the scientific method.
- Branch 1: *Deductive*: e.g. mathematics
- Branch 2: *Empirical*: e.g. statistical data analysis of controlled experiments.

Scientific Discourse

- From an earlier age: print publishing.
- Scientists generally not availing themselves of communication technologies.

Credibility Crisis

- Typical communication doesn't include code, data.
- Published computational science near impossible to replicate.
- Scientific method: Replicability necessary before a discovery is accepted as a contribution to the stock of knowledge. Error control.

Solution: Reproducibility

- New standard for Computational Science needed: release of entire research compendium.
- IP law is a block.

Copyright vs Science

Scientific norms:

1. Copy/replicate results,
2. Build on these results for new discoveries.

=> Communality of research, in exchange for citation.

Copyright vs Science

Copyright establishes property rights in creative works:

- Control over copying
- Control over derivative works

Counter to fundamental scientific norms, yet scientific work falls under copyright by default.

Solution 2: Reproducible Research Standard

Design a mechanism to align copyright
with scientific norms:

- Remove property rights in copying and derivative works.
- Ensure attribution.

Proposal

- Release media components (text, figures) under CC BY.
- Release code components under Modified BSD.
- Both licenses rescind copyright and have an attribution component.

Data

- Raw facts not copyrightable.
- Original “selection and arrangement” of these facts is copyrightable. (Feist Publ’ns Inc. v. Rural Tel. Serv. Co., 499 U.S. 340 (1991))

Principle of Scientific Licensing

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

Share Alike

- Science creates public knowledge.
- License incompatibilities.
- Fencing of the science commons?

- Minimum bar to the use/reuse of scientific knowledge.

Attribution

- Legal attribution and academic citation not isomorphic.
- Minimize administrative burden on scientists.
- Evolving norms / field specific norms.
- Evolving technology.
- “keep intact all copyright notices for the Work and provide, reasonable to the medium or means You are utilizing... .”
- Machine readability.

Why not the Public Domain?

- Attribution incentives.
- Still need to mark.

Benefit for Scientists

- Openness means increased citation.
- Working reproducibly engenders better science.
- Easier for the scientists to build on his or her own work.
- Showcase of skillset for potential collaborators/funders/employers.

But...



"Behind one door is tenure - behind the other is flipping burgers at McDonald's."

Institutional change

- Tenure committees, Presidents
- Journals, peer review
- Funders:
 - RRS provides a hook for funders.
- Communication to researchers.