Intellectual Property Issues in Publishing, Sharing, and Blogging Science

Victoria Stodden
Information Society Project,
Yale Law School
vcs@stanford.edu

ScienceOnline2010
January 16, 2010
Barriers to Sharing 1: Copyright

- Original expression of ideas falls under copyright by default
- Copyright creates exclusive right of the author to:
  - reproduce the work
  - prepare derivative works based upon the original
  - with exceptions: fair use, federal government works, author-designated limitations
This applies to Science..

• Copyright falls automatically on original expression: “follows the ink across the page.”

• E.g. papers, code, blogs, comments, some database structures..
Releasing Data?

- Raw facts alone generally 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))
Creative Commons

• Founded by Larry Lessig to make it easier for artists to share and use creative works

• A suite of licenses that allows the author to determine terms of use attached to works
Creative Commons Licenses

• A notice posted by the author removing the default rights conferred by copyright and adding a selection of:
  • BY: if you use the work attribution must be provided,
  • NC: work cannot be used for commercial purposes,
  • ND: derivative works not permitted,
  • SA: derivative works must carry the same license as the original work.
Open Source Software Licensing

- Creative Commons follows the licensing approach used for open source software, but adapted for creative works.
- Code licenses:
  - BSD license: attribution
  - GNU GPL: attribution and share alike
  - Hundreds of software licenses..
Apply to Scientific Work?

- Remove copyright’s block to fully reproducible research
- Attach a license with an attribution component to all elements of the research compendium (including code, data), encouraging full release.

Solution: Reproducible Research Standard
Reproducible Research Standard

Realignment of legal rights with scientific norms:

- Release media components (text, figures) under CC BY.
- Release code components under Modified BSD or similar.
- Both licenses free the scientific work of copying and reuse restrictions and have an attribution component.
Benefits of RRS

- Focus becomes release of the entire research compendium,
- Hook for funders, journals, universities,
- Standardization avoids license incompatibilities,
- Clarity of rights (beyond Fair Use),
- IP framework supports scientific norms,
- Facilitation of research, thus citation, discovery...
Barriers to Sharing 2: Patents

- Not by default, but by application.
- Since 1980 Bayh-Dole Act encourages technology patenting of federally-funded research by universities to promote tech-transfer to industry.
- Patents disclose underlying discovery but restrict its use through licensing.
- Novel, non-obvious, useful.
- Typically 20 years in duration.
Trademark Law

- Scientists’ work can be associated with scientific output - attribution; identification of unique source.
- Potential corollary of the RRS: branding of work.
Papers

“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”

http://www.stanford.edu/~vcs
Reproducibility

• (Simple) definition: A result is reproducible if a member of the field can independently verify the result.
• Typically this means providing the original code and data, but does not imply access to proprietary software such as MATLAB, or specialized equipment or computing power.
Survey of Computational Scientists

- **Subfield**: Machine Learning
- **Sample**: American academics registered at top Machine Learning conference (NIPS).
- **Respondents**: 134 responses from 638 requests.
# Top Reasons Not to Share

<table>
<thead>
<tr>
<th>Code</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>77%</td>
<td>Time to document and clean up</td>
</tr>
<tr>
<td>44%</td>
<td>Not receiving attribution</td>
</tr>
<tr>
<td>40%</td>
<td>Possibility of patents</td>
</tr>
<tr>
<td>34%</td>
<td>Legal barriers (ie. copyright)</td>
</tr>
<tr>
<td></td>
<td>Time to verify release with admin</td>
</tr>
<tr>
<td>30%</td>
<td>Potential loss of future publications</td>
</tr>
<tr>
<td>52%</td>
<td>Dealing with questions from users</td>
</tr>
<tr>
<td>30%</td>
<td>Competitors may get an advantage</td>
</tr>
<tr>
<td>20%</td>
<td>Web/Disk space limitations</td>
</tr>
</tbody>
</table>
"Behind one door is tenure - behind the other is flipping burgers at McDonald's."

For example..
## Top Reasons to Share

<table>
<thead>
<tr>
<th>Code</th>
<th>Reason</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>91%</td>
<td>Encourage scientific advancement</td>
<td>81%</td>
</tr>
<tr>
<td>90%</td>
<td>Encourage sharing in others</td>
<td>79%</td>
</tr>
<tr>
<td>86%</td>
<td>Be a good community member</td>
<td>79%</td>
</tr>
<tr>
<td>82%</td>
<td>Set a standard for the field</td>
<td>76%</td>
</tr>
<tr>
<td>85%</td>
<td>Improve the caliber of research</td>
<td>74%</td>
</tr>
<tr>
<td>81%</td>
<td>Get others to work on the problem</td>
<td>79%</td>
</tr>
<tr>
<td>85%</td>
<td>Increase in publicity</td>
<td>73%</td>
</tr>
<tr>
<td>78%</td>
<td>Opportunity for feedback</td>
<td>71%</td>
</tr>
<tr>
<td>71%</td>
<td>Finding collaborators</td>
<td>71%</td>
</tr>
</tbody>
</table>
Have you been scooped?

<table>
<thead>
<tr>
<th>Idea Theft</th>
<th>Count</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>At least one publication scooped</td>
<td>53</td>
<td>0.51</td>
</tr>
<tr>
<td>2 or more scooped</td>
<td>31</td>
<td>0.30</td>
</tr>
<tr>
<td>No ideas stolen</td>
<td>50</td>
<td>0.49</td>
</tr>
</tbody>
</table>
Preliminary Findings

- *Surprise*: Motivated to share by communitarian ideals.
- *Not surprising*: Reasons for not revealing reflect private incentives.
- *Surprise*: Scientists not that worried about being scooped.
- *Surprise*: Scientists quite worried about IP issues.