Do Software Patents Slow Scientific Discovery?
Empirical Evidence of the Impact of the Bayh-Dole Act on Computational Science

Victoria Stodden
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign

Seminar Series
Elon Law School
Oct 27, 2014
1. What makes scientific research scientific?
2. How has patenting interacted with the research process?
3. How has computation interacted with the research process?
4. Conflicts with scientific goals
5. Solutions
Merton’s Scientific Norms (1942)

- **Communalism**: scientific results are the common property of the community
- **Universalism**: all scientists can contribute to science regardless of race, nationality, culture, or gender
- **Disinterestedness**: act for the benefit of a common scientific enterprise, rather than for personal gain.
- **Originality**: scientific claims contribute something new
- **Skepticism**: scientific claims must be exposed to critical scrutiny before being accepted
Merton’s Scientific Norms (1942)

- **Communalism**: scientific results are the common property of the community
- **Universalism**: all scientists can contribute to science regardless of race, nationality, culture, or gender
- **Disinterestedness**: act for the benefit of a common scientific enterprise, rather than for personal gain.
- **Originality**: scientific claims contribute something new
- **Skepticism**: scientific claims must be exposed to critical scrutiny before being accepted
**Thesis:** Current Intellectual Property Law distorts incentives in the traditional scientific research process. Specifically, patent and copyright law induce secrecy into a purposefully transparent process.

This reduces the social benefit of the scientific enterprise by slowing scientific progress.
Back to Norms..

- *Skepticism* requires that the claim can be independently verified,

- This in turn requires transparency in the communication of the research process.

- Instantiated by Robert Boyle and the Transactions of the Royal Society in the 1660’s.

- Advances in the technology used for scientific discovery have changed how scientists effect reproducibility.
Advances in Technology

1. enormous, and increasing, amounts of data collection:
   • CMS project at LHC: 300 “events” per second, 5.2M seconds of runtime per year, 5MB per event = 780TB/yr => several PB when data processed,
   • Sloan Digital Sky Survey: 9th data release (SDSS-III 2012), 60TB,
   • quantitative revolution in social science due to abundance of social network data (Lazier et al, Science, 2009),
   • NIH Associate Director for Data Science, 2014.

2. computational power: massive simulations of the complete evolution of a physical system, systematically varying parameters,

3. deep intellectual contributions now encoded in software.
The Scientific Method

Traditionally two branches to the scientific method:

- Branch 1 (deductive): mathematics, formal logic,
- Branch 2 (empirical): statistical analysis of controlled experiments.

Now, new branches due to technological changes?

- Branch 3, 4? (computational): large scale simulations / data driven computational science.

Argument: computation presents only a potential third/fourth branch of the scientific method (Stodden et al 2009).
New Paradigms for Discovery?

Modeling and Simulation:
A NIST Multi-Laboratory
Strategic Planning Workshop

Gaithersburg, MD
September 21, 1995

Workshop Overview

The workshop consisted of an introduction; five talks, each followed by a discussion period; and an open discussion session. Capsule versions follow immediately; more substantial summaries follow later.

Jim Blue opened the workshop with brief introductory remarks. He emphasized that the purpose of doing modeling and simulation is to gain understanding and insight. The three benefits are that modeling and simulation can be cheaper, quicker, and better than experimentation alone. It is common now to consider computation as a third branch of science, besides theory and experiment.

“It is common now to consider computation as a third branch of science, besides theory and experiment.”
The Ubiquity of Error

The central motivation for the scientific method is to root out error:

• Deductive branch: the well-defined concept of the proof,

• Empirical branch: the machinery of hypothesis testing, structured communication of methods and protocols.

Computational science as practiced today does not generate reliable knowledge. Instead, “breezy demos” of results, characterized by the inability to reproduce findings.
Credibility Crisis

Science has lost its way, at a big cost to humanity

Researchers are rewarded for splashy findings, not for double-checking accuracy. So many scientists looking for cures to diseases have been building on ideas that aren’t even true.

Reproducibility

Marcia McNutt

Science advances on a foundation of trusted approaches that scientists use to gain confidence in results. But community wide confidence was shaken by reports that a result was not reproducible. Because confidence in results is so important, we are announcing new initiatives.

NIH Tackles Irreproducibility

The federal agency speaks out about how to improve the quality of scientific research.

By Jef Akst | January 28, 2014
Patents in Science

- Distinction: empirical vs computational science (patents not relevant in deductive science)
- Traditionally patents were most relevant in empirical science
- Now, software is a significant component of scientific research.
- Mertonian norm: Communalism (n.b. federal grants)
Bayh-Dole Act (1980)

• Promote the transfer of academic discoveries for commercial development, via licensing of patents (ie. Technology Transfer Offices),

• Bayh-Dole Act gave federal agency grantees and contractors title to government-funded inventions and charged them with using the patent system to aid disclosure and commercialization of the inventions.

• greatest impact in biomedical research collaborations and drug discovery.
Impact of Computation

In the computational sciences, disclosure of data and code are considered essential for reproducibility. Software can be patent-eligible (Bilski v. Kappos 130 S. Ct. 3218 2010), increasing the reach of Bayh-Dole in the sciences.

Universities can claim ownership over software developed in the course of research on this basis and potentially patent then license access to the code.

Hypothesis: The Bayh-Dole Act inhibits reproducibility in the computational sciences, and is a barrier to access to research inventions.
Ownership of Research Codes

Patent and Copyright Agreement for Personnel at Stanford - SU18

I understand that, consistent with applicable laws and regulations, Stanford University is governed in the handling of intellectual property by its official policies titled Inventions, Patents and Licensing and Copyright Policy (both published in the Research Policy Handbook), and I agree to abide by the terms and conditions of those policies, as they may be amended from time to time.

Pursuant to those policies, and in consideration of my employment by Stanford, the receipt of remuneration from Stanford, participation in projects administered by Stanford, access to or use of facilities or resources provided by Stanford and/or other valuable consideration, I hereby agree as follows:

1. I will disclose to Stanford all potentially patentable inventions conceived or first reduced to practice in whole or in part in the course of my University responsibilities or with more than incidental use of University resources. I hereby assign to Stanford all my right, title and interest in such patentable inventions and to execute and deliver all documents and do any and all things necessary and proper on my part to effect such assignment. (See Inventions, Patents and Licensing for further clarification and discussion related to this paragraph.)

2. I am free to place my inventions in the public domain as long as in so doing neither I nor Stanford violates the terms of any agreements that governed the work done.

3. Stanford policy states that all rights in copyright shall remain with the creator unless the work:
   a. is a work-for-hire (and copyright therefore vests in the University under copyright law),
   b. is supported by a direct allocation of funds through the University for the pursuit of a specific project,
   c. is commissioned by the University,
   d. makes significant use of University resources or personnel, or
   e. is otherwise subject to contractual obligations.

I hereby assign or confirm in writing to Stanford all my right, title and interest, including associated copyright, in and to copyrightable materials falling under a) through e), above.

4. I am now under no consulting or other obligations to any third person, organization or corporation in respect to rights in inventions or copyrightable materials which are, or could be reasonably construed to be, in conflict with this agreement.

NOTE: An alternative to this agreement may be appropriate for personnel with a prior existing and conflicting employment agreement that establishes a right to intellectual property in conflict with Stanford policies. Personnel in this situation should contact the Office of the Vice Provost and Dean of Research.

5. I will not enter into any agreement creating copyright or patent obligations in conflict with this agreement.

6. This agreement is effective on the later of July 1, 2011 (on the one hand) or my date of hire, enrollment, or participation in projects administered by Stanford (on the other hand), and is binding on me, my estate, heirs and assigns.

Electronic Signature in AXESS
http://axess.stanford.edu

The signer should make a copy of this agreement for his or her own records, and hereby waives any objection to Stanford’s use of an electronic version of this agreement as a substitute for the original for any legally recognized purpose.

July 2011

Provider: Office of the Vice Provost and Dean of Research, Stanford University
Contact: Assistant Dean of Research
Last updated: July 2011
Disclosure of Research Codes

Claim: Codes would (eventually) be fully open in the absence of Bayh-Dole:

• Grassroots “Reproducible Research” movement in computational science (policy development, best practices, tool development),

• Changes in funding agency requirements

• Changes in journal publication requirements
“Reproducible Research” is Grassroots

• AMP 2011 “Reproducible Research: Tools and Strategies for Scientific Computing”
• Open Science Framework / Reproducibility Project in Psychology
• AMP / ICIAM 2011 “Community Forum on Reproducible Research Policies”
• SIAM Geosciences 2011 “Reproducible and Open Source Software in the Geosciences”
• ENAR International Biometric Society 2011: Panel on Reproducible Research
• AAAS 2011: “The Digitization of Science: Reproducibility and Interdisciplinary Knowledge Transfer”
• SIAM CSE 2011: “Verifiable, Reproducible Computational Science”
• Yale 2009: Roundtable on Data and Code Sharing in the Computational Sciences
• ACM SIGMOD conferences
• NSF/OCI report on Grand Challenge Communities (Dec, 2010)
• IOM “Review of Omics-based Tests for Predicting Patient Outcomes in Clinical Trials”
• ...


Best Practices for Computational Science: Software Infrastructure and Environments for Reproducible and Extensible Research

Victoria Stodden
Columbia University - Department of Statistics
Sheila Miguez
Columbia University
September 6, 2013

Abstract: Scholarly dissemination and communication standards are changing to reflect the increasingly computational nature of scholarly research, primarily to include the sharing of the data and code associated with published results. This paper presents a formalized set of best practice recommendations for computational scientists wishing to disseminate reproducible research, facilitate innovation by enabling data and code re-use, and enable broader communication of the output of digital scientific research. We distinguish two forms of collaboration to motivate choices of software environment for computational scientific research. We also present these Best Practices as a living, evolving, and changing document on wiki.
New Tools for Computational Reproducibility

- Dissemination Platforms:
  - ResearchCompendia.org
  - MLOSS.org
  - Open Science Framework
  - IPOL
  - thedatahub.org
  - The DataVerse Network
  - Madagascar
  - nanoHUB.org
  - RunMyCode.org

- Workflow Tracking and Research Environments:
  - VisTrails
  - Kepler
  - CDE
  - Galaxy
  - GenePattern
  - Synapse
  - Sumatra
  - Taverna
  - Pegasus

- Embedded Publishing:
  - Verifiable Computational Research
  - Sweave
  - knitR
  - Collage Authoring Environment
  - SHARE
Sharing: Funding Agency Policy

- NSF grant guidelines: “NSF ... expects investigators to share with other researchers, at no more than incremental cost and within a reasonable time, the data, samples, physical collections and other supporting materials created or gathered in the course of the work. It also encourages grantees to share software and inventions or otherwise act to make the innovations they embody widely useful and usable.” (2005 and earlier)

- NSF peer-reviewed Data Management Plan (DMP), January 2011.

- NIH (2003): “The NIH expects and supports the timely release and sharing of final research data from NIH-supported studies for use by other researchers.” (>$500,000, include data sharing plan)
National Science Board Report


Open Science from the Whitehouse

• Feb 22, 2013: Executive Memorandum directing federal funding agencies to develop plans for public access to data and publications.

• May 9, 2013: Executive Order directing federal agencies to make their data publicly available.

• July 29, 2014: Notice of Request for Information “Strategy for American Innovation”
Executive Memorandum: “Expanding Public Access to the Results of Federally Funded Research”

- “Access to digital data sets resulting from federally funded research allows companies to focus resources and efforts on understanding and exploiting discoveries.”

- “Digitally formatted scientific data resulting from unclassified research supported wholly or in part by Federal funding should be stored and publicly accessible to search, retrieve, and analyze.”

- “Digital recorded factual material commonly accepted in the scientific community as necessary to validate research findings”

- “Each agency shall submit its draft plan to OSTP within six months of publication of this memorandum.”
Executive Order: “Making Open and Machine Readable the New Default for Government Information”

- “The Director … shall issue an Open Data Policy to advance the management of Government information as an asset”
- “Agencies shall implement the requirements of the Open Data Policy”
- “Within 30 days of the issuance of the Open Data Policy, the CIO and CTO shall publish an open online repository of tools and best practices”
Request for Input: “Strategy for American Innovation”

- “to guide the Administration's efforts to promote lasting economic growth and competitiveness through policies that support transformative American innovation in products, processes, and services and spur new fundamental discoveries that in the long run lead to growing economic prosperity and rising living standards.”

- “(11) Given recent evidence of the irreproducibility of a surprising number of published scientific findings, how can the Federal Government leverage its role as a significant funder of scientific research to most effectively address the problem?”
Science Policy in Congress

- America COMPETES due to be reauthorized, drafting underway.
- Hearing on Research Integrity and Transparency by the House Science, Space, and Technology Committee (March 5, 2013).
- Reproducibility cannot be an unfunded mandate.
NAS Data Sharing Report


• “Principle 1. Authors should include in their publications the data, algorithms, or other information that is central or integral to the publication—that is, whatever is necessary to support the major claims of the paper and would enable one skilled in the art to verify or replicate the claims.”
• March 23 2012, Institute of Medicine releases report,

• Recommends new standards for omics-based tests, including a fixed version of the software, expressly for verification purposes.
“The fully specified computational procedures are locked down in the discovery phase and should remain unchanged in all subsequent development steps.”
### Barriers to Sharing

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>77%</td>
<td>Time to document and clean up</td>
<td>54%</td>
</tr>
<tr>
<td>52%</td>
<td>Dealing with questions from users</td>
<td>34%</td>
</tr>
<tr>
<td>44%</td>
<td>Not receiving attribution</td>
<td>42%</td>
</tr>
<tr>
<td>40%</td>
<td>Possibility of patents</td>
<td>-</td>
</tr>
<tr>
<td>34%</td>
<td>Legal Barriers (i.e. copyright)</td>
<td>41%</td>
</tr>
<tr>
<td>-</td>
<td>Time to verify release with admin</td>
<td>38%</td>
</tr>
<tr>
<td>30%</td>
<td>Potential loss of future publications</td>
<td>35%</td>
</tr>
<tr>
<td>30%</td>
<td>Competitors may get an advantage</td>
<td>33%</td>
</tr>
<tr>
<td>20%</td>
<td>Web/disk space limitations</td>
<td>29%</td>
</tr>
</tbody>
</table>

Survey of the Machine Learning Community, NIPS (Stodden 2010)
Back to Bayh-Dole

Potential implications for science as a disruptor of openness norms:

• delay in revealing code, or closed code,

• potential obfuscation of methods submitted for patents (Bilski v. Kappos),

• alteration of a scientist’s incentives toward commercial ends, instead of the production of science as a public good.
# Classifying Software Patents

<table>
<thead>
<tr>
<th>PTO Classification Code</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>341</td>
<td>Coded Data Generation or Conversion</td>
</tr>
<tr>
<td>345</td>
<td>Computer Graphics Processing</td>
</tr>
<tr>
<td>370</td>
<td>Multiplex Communications</td>
</tr>
<tr>
<td>706</td>
<td>Data Processing: Artificial Intelligence</td>
</tr>
<tr>
<td>707</td>
<td>Data Processing: Database and File Management or Data Structures</td>
</tr>
<tr>
<td>708</td>
<td>Electrical Computers: Arithmetic Processing and Calculating</td>
</tr>
<tr>
<td>716</td>
<td>Computer-aided Design and Analysis of Circuits and Semiconductor Masks</td>
</tr>
<tr>
<td>717</td>
<td>Data Processing: Software Development, Installation, and Management</td>
</tr>
</tbody>
</table>
Academic Software Patenting

Total Number of Software Patents filed by the top 23 University Patent Filers, 2000-2012

Software Patents as a Percent of the Total University Patent Portfolio, 2000-2012
Modeling Patent Seeking

We postulate:

• a positive relationship between patenting and industry collaboration, and

• a positive relationship between academic publication and patenting behavior.
Simple Model

\[ Y_t = \beta_t + \text{Ind.Collab}_t + \text{Ac.Pub.}2\text{to}5_t + \text{Ac.Pub.}1_t + t + \epsilon_t, \quad t = 2000 \ldots 2012, \]

- \( Y_t \): number of software patents filed by the 23 universities or colleges indexed by year,
- \( \text{Ind.Collab}_t \): the number of patents with an industry researcher as co-author,
- \( \text{Ac.Pub.}2\text{to}5_t \): counts the number of patents that refer to an academic article within 2 to 5 years of the data of filing,
- \( \text{Ac.Pub.}1_t \): whether there was a referenced academic publication within a year of the patent filing date (i.e., exempt from being classified as prior art if they are within one year of application date).
Model Coefficients for Academic Software Growth

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Coefficient Estimate</th>
<th>Standard Error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta )</td>
<td>36.78</td>
<td>9.71</td>
<td>0.0053</td>
</tr>
<tr>
<td>Ind.Collab</td>
<td>7.34</td>
<td>2.64</td>
<td>0.0239</td>
</tr>
<tr>
<td>Ac.Pub.2to5</td>
<td>-3.36</td>
<td>2.61</td>
<td>0.2342</td>
</tr>
<tr>
<td>Ac.Pub.1</td>
<td>0.12</td>
<td>1.79</td>
<td>0.9478</td>
</tr>
<tr>
<td>t</td>
<td>4.47</td>
<td>1.10</td>
<td>0.0036</td>
</tr>
</tbody>
</table>

- Industry collaboration indicates higher likelihood of patent seeking,
- No other significant relationships (except time i.e. omitted variables)
Model Conclusions

- Software patents not a significant part of university patent portfolios, but growing,

- little pressure not to patent from other aspects of the research process (i.e. publication requirements, TTO pressure), but changing,

- Note 1: institutional pressure to patent exists.

- Note 2: omitted variables such as, overall publication counts by author, startup counts, institutional policies, ...
Conclusions

• Incentives and requirements to patent create a competing route to software transparency, over open release.

• This route can provide renumeration to institutions, who hold the rights to the research inventions (Bayh-Dole).

• Claim 1: Incentives to patent are siphoning software from open release into licensed access.

• Claim 2: As pressure for open code in research increases, patenting will come in conflict with transparency.

• Claim 3: Open code is central to reproducibility in computational science, and limiting access impedes the production of reliable scientific conclusions.
Workaround 1: Dual Licensing

• Distinguish between industry and academic research applications for licensing,

• Standard for academic research, in particular results associated with published results, and publicly funded research,

• Code for academic use simple to download, with associated licensing terms (no interaction with TTO).
Workaround 2: Public Access

- Increase the role and responsibilities of the Technology Transfer Office to a dual mandate (Stodden, Rules for Growth, 2011):
  1. Shepherd patents and licensing agreements,
  2. Shepherd public access to digital scholarly objects.
- Increase the role and responsibilities of institutional libraries:
  - curation, archiving, persistent access to digital scholarly objects.
Other Legal Barriers to Open Code

- HIPAA (Health Information Portability and Accountability Act) and privacy regulations,
- Copyright (i.e. Reproducible Research Standard),
- Collaboration agreements with industry,
- Hiring agreements, institutional rules,
- National security.
Selected References

• Reproducible Research, Guest editor for Computing in Science and Engineering, July/August 2012.


• “Reproducible Research: Tools and Strategies for Scientific Computing,” July 2011

available at http://www.stodden.net