

# Opening the Black Box: Sharing in the 'Republic of Science'

Victoria Stodden

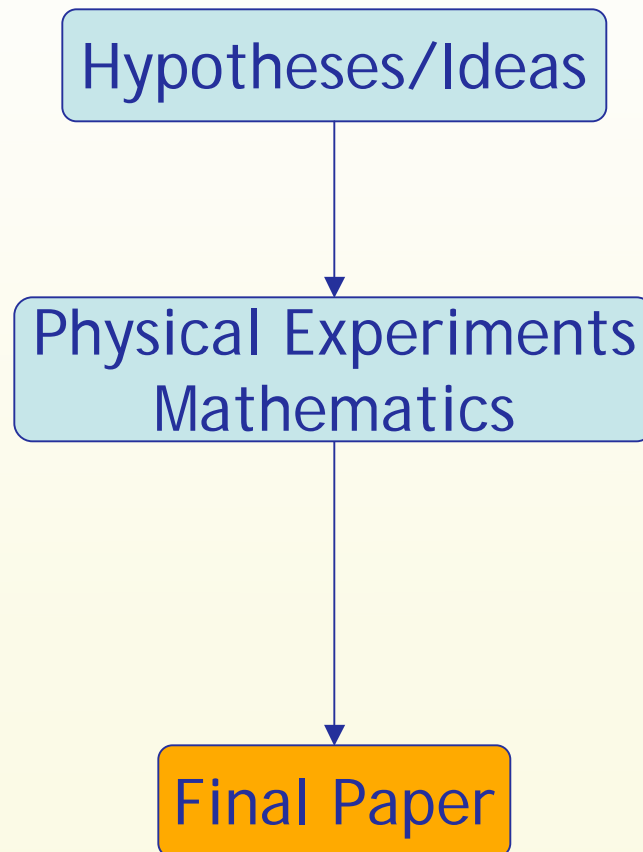
[vcs@stanford.edu](mailto:vcs@stanford.edu)

*User and Open Innovation Workshop*

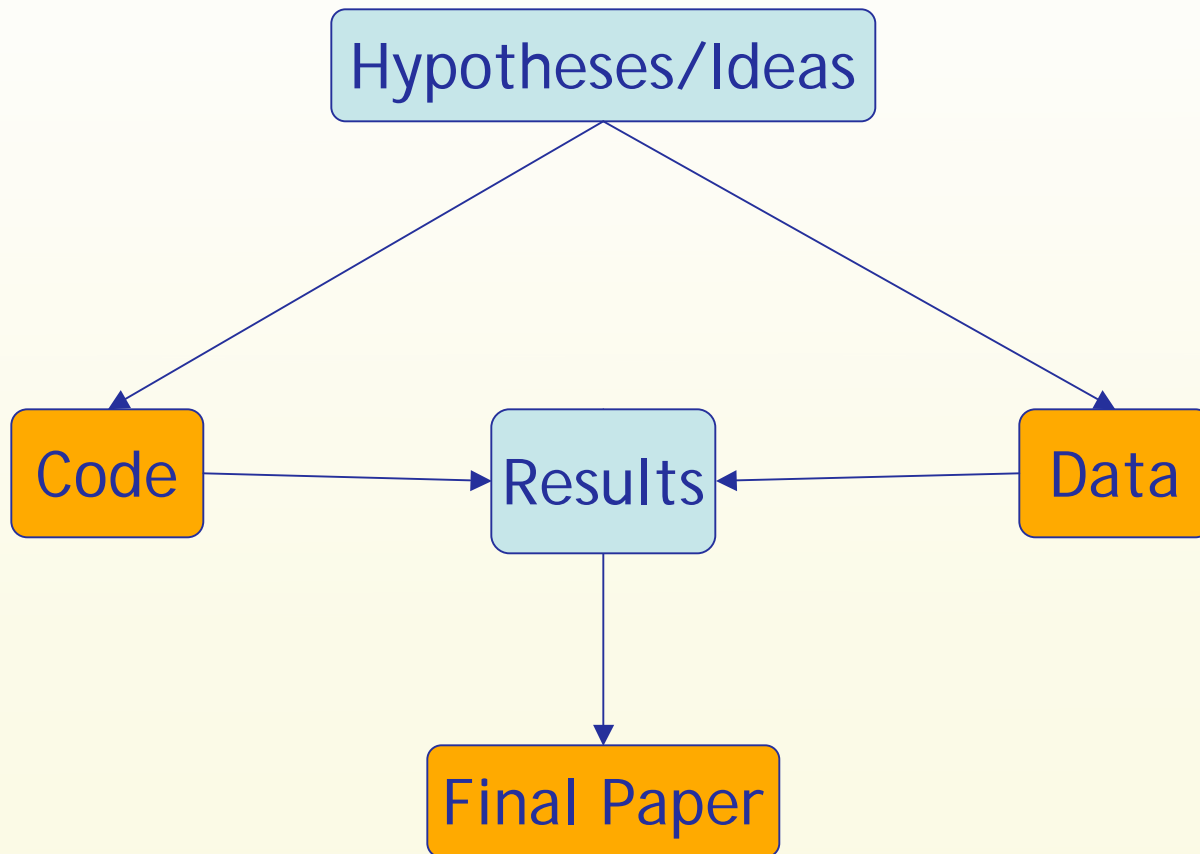
Hamburg University of Technology, Germany

June 4, 2009

# Scientific Output is Changing



# Scientific Output is Changing

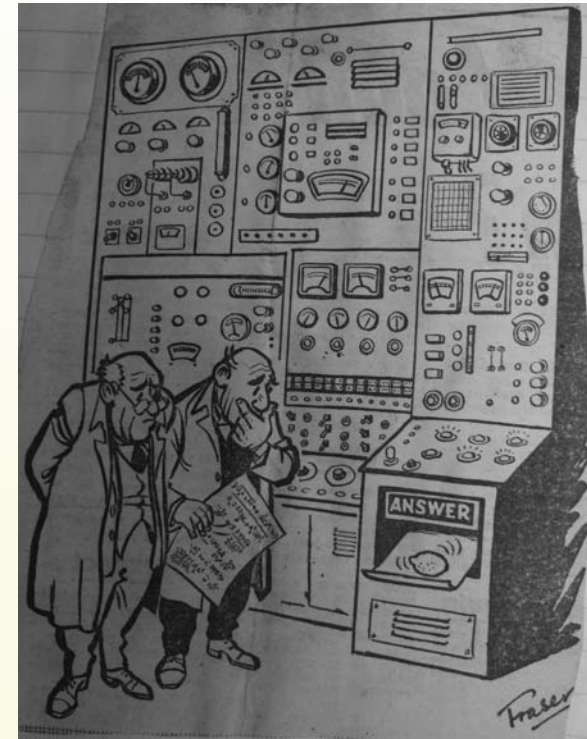


# Computation is Increasingly Pervasive

- JASA June 1996: 9 of 20 articles computational (none made code or data available).
- JASA June 2006: 33 of 35 articles computational (3 had code publicly available, no data).

# Framework for Scientific Revealing

- Crisis in computational research: results not reproducible.
- When do scientists reveal and when do they conceal? Are there infant commons?
- Literature on Free Revealing in industry and Sociology of Science



# Hypotheses for Revealing

- Scientists perceive *personal* gain:
  - reputation gains,
  - claims on priority,
  - black box closure,
  - acceleration of the research process,
  - advantage through the establishment of standards...
- The willingness to reveal work reflects the *desire for community membership and feedback*.
- Free riders do not gain as much as revealers in their use of the revealer's work.

# Survey of Computational Scientists

- *Subfield*: Machine Learning.
- *Sample*: American academics registered at the top Machine Learning conference (NIPS).
- Many fields represented: mathematics, statistics, psychology, neurology, computer science, electrical engineering, linguistics, robotics, medicine...
- *Respondents*: 130 responses from 595 requests (excluding 43 bounces). 22%

# Top Reasons Not to Share

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



For example..



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

# Top Reasons to Share

<i>Code</i>		<i>Data</i>
91%	Encourage scientific advancement	81%
90%	Encourage sharing in others	79%
86%	Be a good community member	79%
82%	Set a standard for the field	76%
85%	Improve the caliber of research	74%
81%	Get others to work on the problem	79%
85%	Increase in publicity	73%
78%	Opportunity for feedback	71%
71%	Finding collaborators	71%

# Preliminary Findings

- *Surprise*: Scientists are generally motivated to share by communitarian ideals.
- *Not surprising*: Reasons for not revealing reflect private incentives ie. opportunity for feedback and community membership.
- *Surprise*: Scientists not that worried about being scooped. Free riding tolerated?
- Legal barriers a factor: Reproducible Research Standard

# To Come..

- Document respondents' characteristics and actual sharing behavior on the web - correlate with survey results.
- Prediction of sharing behavior.
- Delineation of non-response bias and response bias.
- Granularity of knowledge divisions.