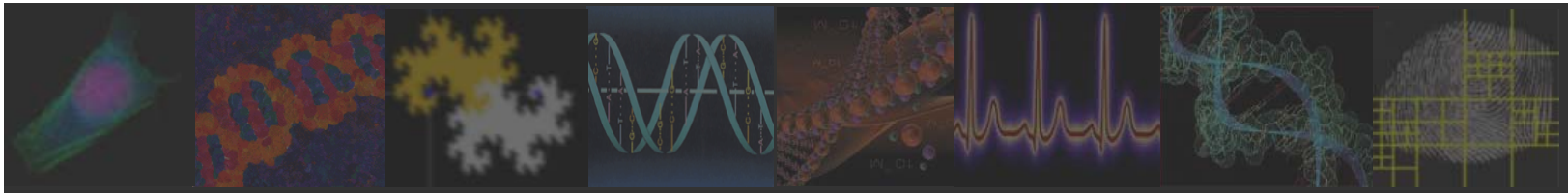
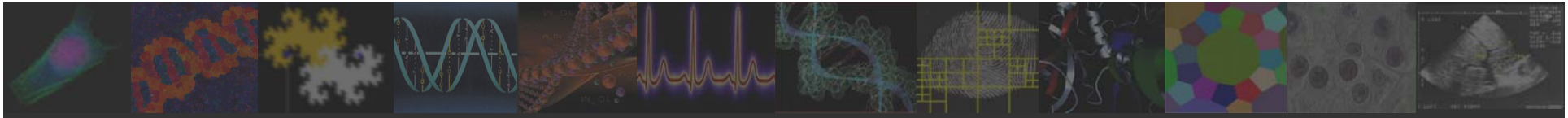


How to Encourage and Publish Reproducible Research

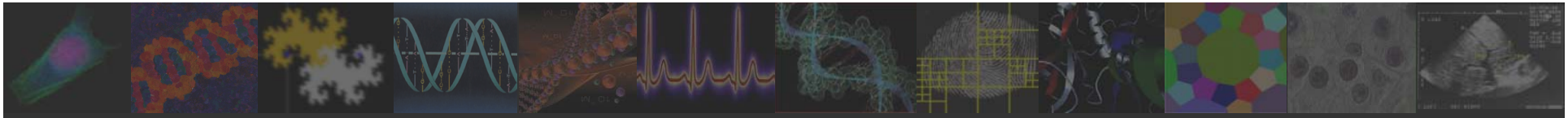
Jelena Kovačević

Center for Bioimage Informatics
Department of Biomedical Engineering
Department of Electrical & Computer Engineering
Carnegie Mellon University



Outline

- What is RR and why do we need it?
  - Theory versus experimentation
  - A hybrid is born: Computational sciences
  - Birth of RR
  - Why do we need RR?
- How do we get to RR?
  - Issues to consider
    - Cultural, educational, data, IP
  - Suggested course of action
  - How do we publish RR?
    - How to write RR papers and tools to enable RR
- An entirely nonRR case study

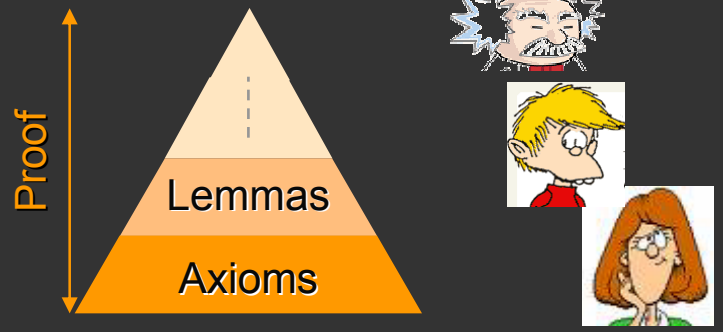


What is RR and Why Do We Need It?

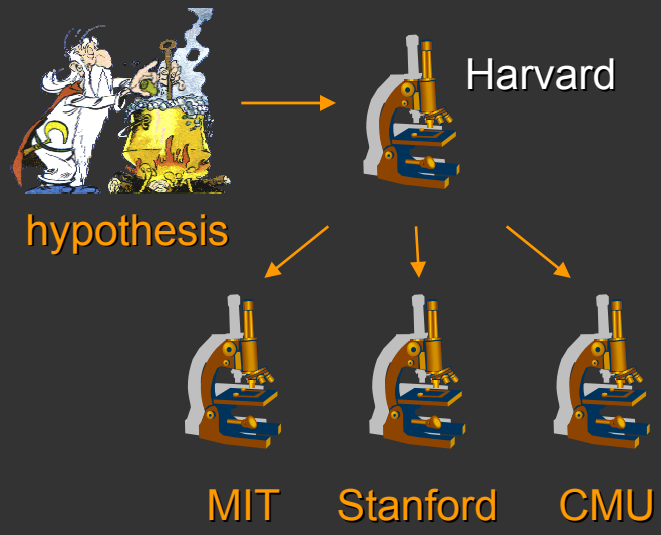
- 1993: Cracking math's oldest brain-teaser
  - Mathematics
  - Proof of Andrew Wiles of Fermat's Last Theorem
  - Took mathematicians and Wiles 2 years to prove/check
  - RR: The proof was "reproduced"/validated by others
- Cloned human embryos are stem cell breakthrough
  - Biology
  - Suk scandal
  - Cell lines were doctored/scientist coerced into donating eggs
  - RR: The results could not be reproduced
- RR
  - Concept: in "computational" sciences, the ultimate product is not a published paper but rather the entire environment used to produce the results in the paper (data, software, etc.).
  - Natural and obvious: how many of us really do it that way?
  - WHY do we need it?

Theory Vs Experimentation

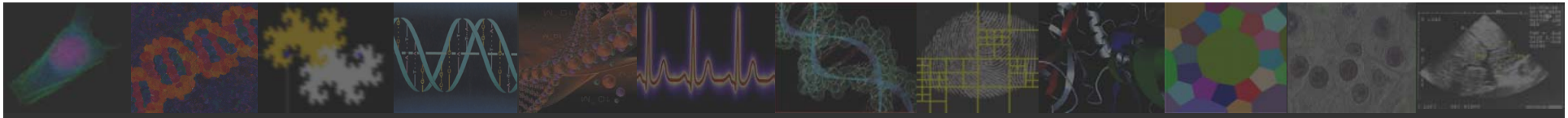
- Theoretical disciplines
  - Mathematics



- Experimental disciplines
  - Biology

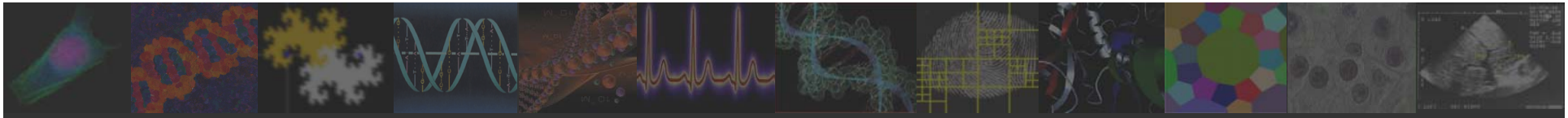


- A hybrid is born: Computational sciences
  - Should follow good practices from both
  - SP falls in there: How are we doing?



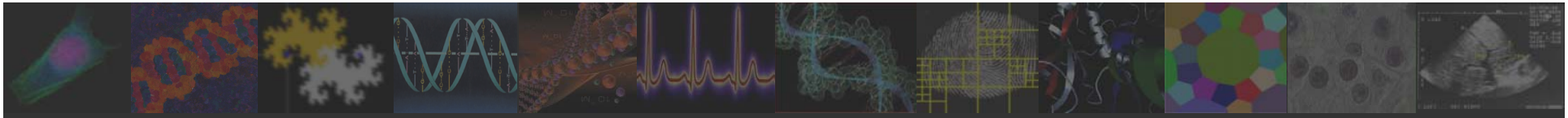
Birth of RR

1984	Knuth	Literate programming • Programs useless without descriptions • Extract code from descriptions
early 1990s	Claerbout, Pouzat	Article only advertisement of scholarship • Real scholarship: data and software
1995	Buckheit & Donoho	Closer to our area
2000s	Greyer	Requirements for RR
2005	Barni & Perez-Gonzales	Opinion piece in SP Magazine
2000s	Vetterli Vandewalle	Promoting at the EPFL site



Issues

- Cultural
  - Innovation above all else
  - TIP Transactions reviewing questions
    - 1. Is the paper technically sound?
    - 2. Is the coverage of the topic sufficiently comprehensive and balanced?
    - 3. How would you describe the technical depth of the paper?
    - 4. How would you rate the technical novelty of the paper?
  - Can lead to paradox
- Educational
  - Our students undertrained in statistics
  - Typically reimplement everything
- Data
  - We collaborate and data might not be ours
- IP
  - Data issues
  - Companies and agencies protecting their IP
- Collaborative
  - With colleagues within the university/company, outside

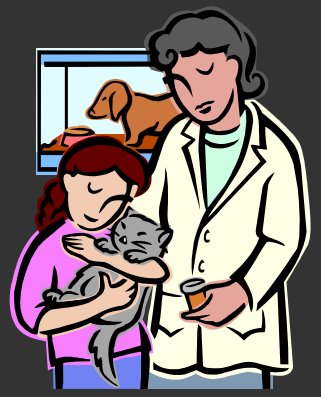


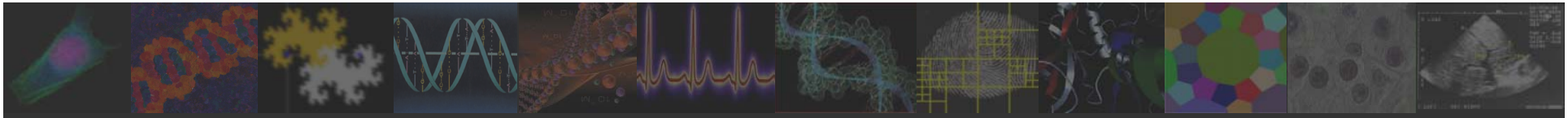
Suggested Course of Action

- Encourage authors to publish first-class, experimental work.
- Encourage authors to submit work which uses a known algorithm in a new setting or with a different type of data.
- Show value of such work by publishing special issues, promoting it through paper awards and training students to perform such work.
- Blueprint for papers accepted for publication
  - Code, software, readme file, …
- Negotiate for a representative data sample to be available when data is protected.
- Promote the idea of RR with the national funding agencies.
- Develop templates of what should be published and how. Develop templates for collaborative work and sharing of data.

How Do We Publish RR?

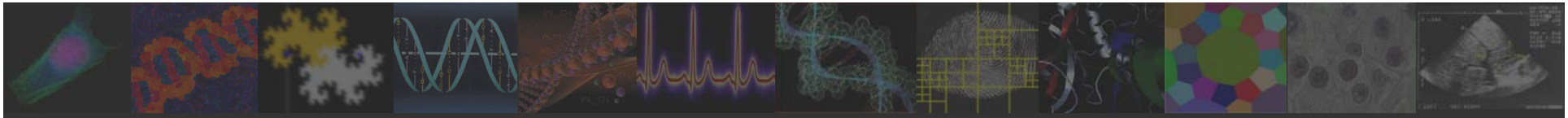
- Not likely to happen overnight
  - Encourage and reward “good behavior” (Child psychology 101)
- Ideas
  - Special section in Transactions for RR?
  - Establish a paper award for an RR paper?
  - Form a rough guideline of what each paper should contain for an RR designation?
  - Everything we read is partly “on faith”





How to Write and Make Papers RR?

- Example
  - Used in my group
  - Compilation of ideas from Barni and EPFL groups (Vetterli, Vandewalle et al.)
  - Compendium (Gentleman & Lang)
    - Freeze the code upon
      - Submission
      - Acceptance
  - “Good intentions” (Marziliano) enforced
  - Students do projects and reproduce



An Entirely NonRR Case Study

- Data set
  - 15 papers published in the TIP
  - EDICS category using both theory and experimentation
  - Stayed away from standards as well as biomed
  - For all algorithms, competing ones exist
- Ratings (0, 0.5, 1)
  - Algorithm and experimental setup
    - algorithm explained?
    - data explained?
    - data size?
    - details on parameters used?
    - comparison to competing algorithms?
  - RR
    - block-diagram?
    - pseudo code?
    - data available?
    - code available?
    - proof available?

Results of the Entirely NonRR Case Study

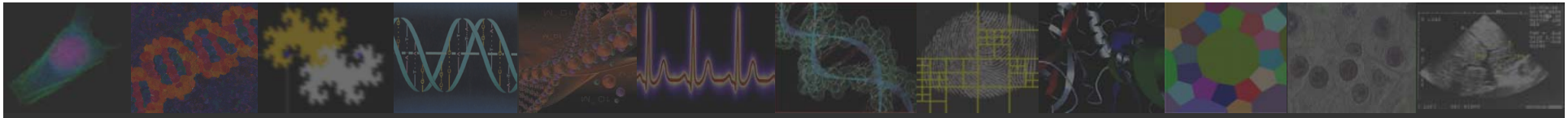
Algorithm and Experimental Setup [%]				
Algorithm details	Data details	Data size	Parameter details	Comparisons
80	33	46	46	26
Reproducible Research Criteria [%]				
Block diagram	Pseudo code	Data available	Code available	Proof available
0	60	33	0	100

- All papers had proofs, none had code available
- Sufficient detail on algorithms, none had a block-diagram
- Data used, data size and availability all below average
- Half of the cases were the parameters specified
- Comparisons to competing algorithms: quarter
- Pleasant surprise: 60%, pseudo-code was available

Results of the Entirely NonRR Case Study

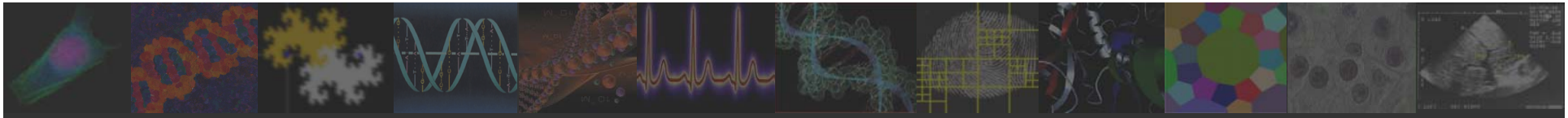
- How Did I Do?
  - Algorithm and experimental setup
    - algorithm explained? 1
    - data explained? 0.5
    - data size? 0.5
    - details on parameters used? 0
    - comparison to competing algorithms? NA
  - RR
    - block-diagram of the algorithm? 0
    - pseudo code of the algorithm? 0
    - data available? 0
    - code available? 0
    - proof available? NA

- So you are left to believe me when I give you the above numbers.
  - Should you? Of course not!



Can We Make It Happen?

YES!



Acknowledgments and Disclaimers

- Thanks!
  - Mauro Barni and Fernando Perez-Gonzalez
  - Martin Vetterli
  - Patrick Vandewalle
  - Members of the TIP Editorial Board
  - Informal email group Mauro and Fernando organized
- Thoughts expressed
  - When not cited, my opinions on the issue