Panel: *"Reproducibility in Visualization"* or *"Do we want a Badge of Honor?"*

Rosane Minghim

ICMC – University of São Paulo Brazil

Focusing on some of the Issues

Reproducibility

- Code and data preparation
- Intellectual property

Validation

- Measuring the visual

" The process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model." (AIAA G-077-1998)

Some more on (precious?) code
Reasons not to share

- **Intellectual property issues**
  - " The code is valuable intellectual property that belongs to my institution.

  Really, that little MATLAB routine to calculate a two-part fit is worth money? Frankly, I doubt it. Some code may have long-term commercial potential, but almost all the value lies in your expertise. My industry has a name for code not backed by skilled experts: abandonware. Institutions should support publishing; those who refuse are blocking progress." *
  - Change of policy from sponsors.

- **Issues of good standards for code**
  - "It is too much work to polish the code.

  For scientists, the word publication is totemic, and signifies perfectionism. […] the original code can be published as supplementary information, available from an institutional or journal website." *
  - There is no obligation to produce quality code.

** Nick Barnes, Nature 467, 753 (2010)*
Publish your code, it is good enough

Some more on (precious?) code
Reasons to share

- **Academic Reasons**
  - Enables others to engage in your research. *
  - Improves the chance of impact of your research.
  - Leads to people referencing your work.
    Maybe being referenced is a more genuine measure than number of papers?
    What are the most referenced papers on techniques?

- **Development Reasons**
  - "Final products are not all that polished anyway". *
  - Starting from code and data is better that starting from words.
  - Allowing your techniques to be improved and adapted by others may lead to reaching higher goals.
  - Promoting your ideas and techniques helps them to become standard procedure faster.

  After all, some of us probably got into this business to improving people´s lives and have our ideas change how people work on theirs.
  - Making available is the next best choice to helping people to code. There are well intentioned non-profit groups in need of code.

** Nick Barnes, **Nature** 467, 753 (2010)*
Publish your code, it is good enough

Some more on Data

- **Data is result of hard and delicate work**
  - Biology, astrophysics, and so many other specialties leave on producing data.
  - Privacy, security, sensitivity issues.
  - Data is disguised due to IP (again), privacy or sensitivity issues.
    - At least make the disguised data available.
- **Some things are changing…**
  - Much more available data every time in various applications (ex. climate, news, simulations, etc.).
  - Reproducibility standards are being supported more by various venues.
  - Publishing data may be a form of publication – more on that next.

Inspiration from other Fields

Imaging, data mining, computational applications

- Ex. Journals that encourage reproducibility as available code and data (see *).
  Biostatistics, Oxford Journals
  (Impact factor 2.769)

- Ex. papers on benchmarks

P. Dollár, C. Wojek, B. Schiele and P. Perona
Pedestrian Detection: An Evaluation of the State of the Art
PAMI, 2011. Impact Factor 4.38.

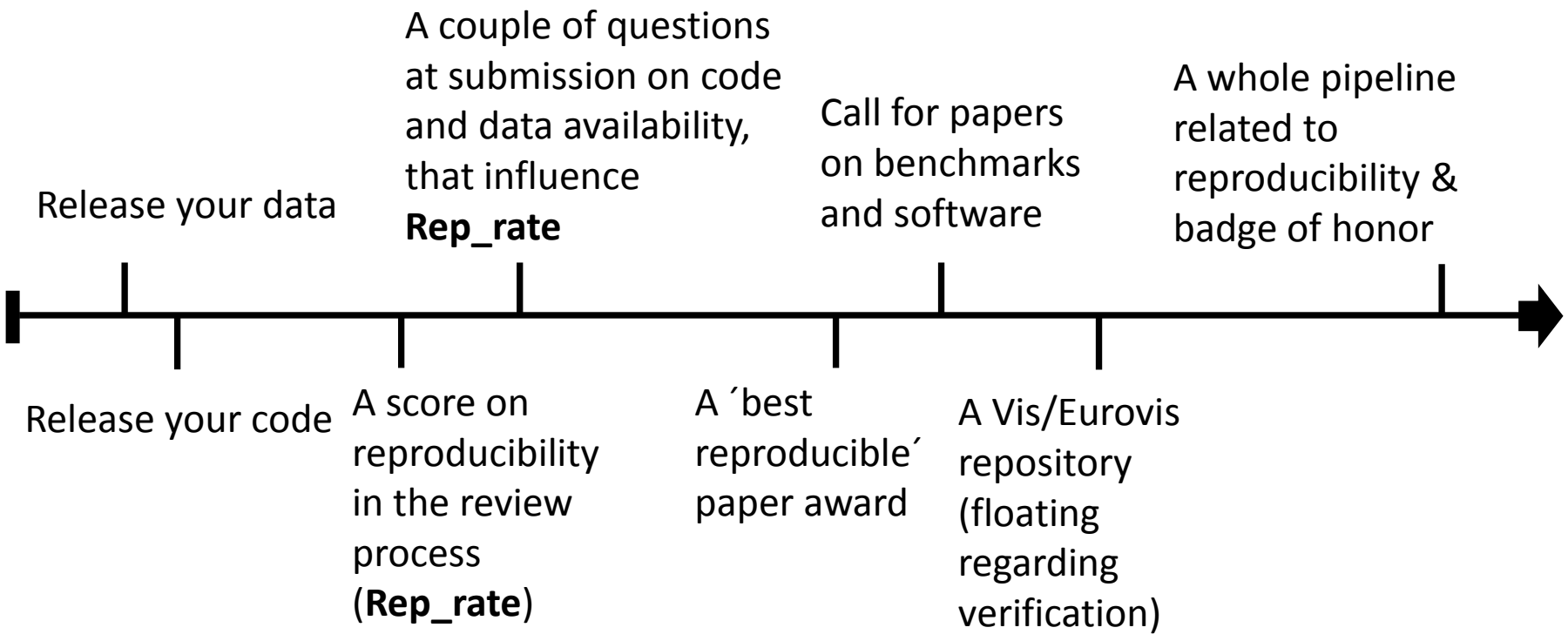
P. Dollár, C. Wojek, B. Schiele and P. Perona
Pedestrian Detection: A Benchmark
CVPR 2009.

- Ex. Available benchmarks in imaging
  Caltech Pedestrian Detection Benchmark:
  http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/

  ETHZ: Computer Vision Lab (Datasets)
  http://www.vision.ee.ethz.ch/datasets/index.en.html

** D. L. Donoho, D. L., **Biostatistics** 11,3 (2010)*
An invitation to reproducible computational research

Thinking of solutions: degrees of commitment to reproducibility.



On Validation and Verification in Vis papers

- Must verify
- Should compare
- Should validate
- Should divulge the parameters!!

- There are numbers to validate visualizations:
  - SciVis

    eg. Information Theory measurements, mesh quality measurements, etc.
  - Infovis/Visual Mining

    Stress, distance plots, neighborhood measurements , silhouette, correlation coefficient, etc.

  So one should use them.

- Perceptual Evaluation is called for as separate studies, helping verify correspondence with the numbers and with claims.

Having said that…

- Encouragement to release code and data should not restrict initiatives.
- High degree of validation or verification can be set aside on occasion .
  - An interesting idea should not be left behind.
- Validation may not even be possible when there is dilemma/uncertainty involved → precision values cannot be used

Thanks to

Rodrigo de Lima Pinheiro

Alejandro Frery

Hélio Pedrini

Alex Telea

Anton Heijs

and of course…

Britta Weber

Paul Rosenthal