Experiences with Reproducible Research in Various Facets of Signal Processing Research

P. Vandewalle, G. Barrenetxea, I. Jovanovic, A. Ridolfi and M. Vetterli
Ecole Polytechnique Fédérale de Lausanne (EPFL)

April 18, 2007
**Introduction**

**Definition:**

“Reproducibility is one of the main principles of the scientific method, and refers to the ability of a test or experiment to be accurately reproduced, or replicated, by someone else working independently.”

- Wikipedia -

Different parts needed for reproducibility:

- Publication
- Data
- Code

**Advantages:**

- Allows you to continue where you left work
- Allows others to start from the same point
- Allows others to use your work as a building block
Introduction

- own activity
  - new results
  - writing paper
  - paper revision
  - picking up previous work

- others’ activity
  - citations
  - industrial applications
  - teaching

$t_0$
Introduction

Increase impact!
- higher level
- faster

own activity

others’ activity

citations
industrial applications
teaching
Introduction

Some examples:

“I just read your paper X. It is very completely described, however I am confused by Y. Could you provide the implementation code to me for reference if possible?”

“Comment from a reviewer about a paper: Please replace Figure 3 by a higher quality version. Also, the horizontal axis should be labeled ‘t’ instead of ‘x’.”

“Hi! I am also working on a project related to X. I have implemented your algorithm, but cannot get the same results as described in your paper. Which values should I use for parameters Y and Z?”
Related Work - Other Domains

- Mathematics
  - ‘Reprovable’, proofs can be verified
  - Not reproducible: Fermat’s last theorem
    “J'ai trouvé une merveilleuse démonstration de cette proposition, mais je ne peux l'écrire dans cette marge car elle est trop longue.”

- Exact Sciences (Physics, Chemistry, etc.)
  - Experimental setups

- Life Sciences
  - Experimental setups
  - Other researchers repeat experiments
  - Journal of Cell Biology checks for image manipulation

→ We have a lot to learn!
Related Work - History of Reproducible Research

• Knuth (1984)
  – Literate Programming
    “Instead of imagining that our task is to instruct a computer what to do, let us concentrate rather on explaining to human beings what we want a computer to do.”
  
• Claerbout @ SEP, Stanford (1990)
  – Makefiles to build and clean results

• Donoho @ Statistics Department, Stanford (1995)
  – WaveLab using Matlab

• Various other domains
  – Econometrics [Koenker96, Vinod01]
  – Neurophysiology [Pouzat05]
  – Epidemiology [Peng06]
  – Signal Processing

→ Now is the right time to start Reproducible Research!
Parts of Reproducible Research - Theory

Theory
- Well described theorems and proofs
- Numerical simulations are very helpful

Example: Spectrum of Ultrawide Bandwidth Signal

\[ S_X(\nu) = |\tilde{\omega}(\nu)|^2 \lambda \left( 2 \text{Re} \left\{ \sum_{k \geq 0} \phi_S^k (2\pi \nu) \right\} d\nu - 1 - \lambda \delta(\nu) \right). \]

Parts of Reproducible Research - Algorithms

Algorithms

• Code
• Data
• Environment
  – Computer platform
  – Compiler and compiler flags
  – Software version
  – User interface, if available
Data and Experimental Setups

- Measurement setup
- Setup or calibration procedures
- Complex setups: reusable data sets

Examples: SensorScope, Acoustic Tomography


Example: Super-Resolution Imaging

- First and second conference paper - 2003
  - Not reproducible
- First journal paper - April 2004
  - Let’s make it reproducible…
  - 1 week work
  - Paper with Matlab code, data, figures online
- …
- Second journal paper - Dec 2005
  - Reproducible, of course!
  - 1 day work
  - Paper with Matlab code, data, and figures
- In between (summer 2005), student wrote GUI for Matlab code
- Writing PhD thesis
  - Great, I can recycle my code!
  - Easy to create new figures by changing parameters and axis labels
  - 3 months to write thesis
Example: Super-Resolution Imaging

Web page for each paper, containing:

- Full reference + PDF
- Abstract
- Code
- Data
- Tested configurations
- Contact address
- (Figures)
- (References)
Example: Super-Resolution Imaging

Benefits:
• Efficient reuse of my own results
• Nice demo material (using GUI)
• Many downloads

• Nice reactions
  – “When I am beginning to study image super-resolution, I read a lot of papers about it, but I found that I don't know how to realize the algorithms of these paper. Your works help me make great progress. Hope to have more super-resolution algorithms code for us to study. Thanks.”
  – “We're interested in improving the algorithm via our own methodology in registration. We're using your methods as a springboard of research for use in security applications.”

• Collaborations
Reproducible Paper Repository

How to make those data available?

• Everyone creating own web pages with more/less information
• Centralized repository with reproducible papers (cfr ArXiv)
• Distributed repository setup with standardized fields
  – Using EPrints server (www.eprints.org) and configuration files
  – Every lab/institution can build own database
  – Compatibility
  – We are currently building a setup jointly with EPrints
Licensing and Commercialization

- License defines how your code can be used and distributed
- License has to be included with the code
- Many types of licenses exist
  - Open Source
    - GNU General Public License (GPL)
    - BSD License
    - MIT License
    - Common Public License (CPL)
    - … (58 listed currently on http://www.opensource.org/licenses/)
  - Proprietary Source

What if my start-up has to live from this?
- Special setup: website where people can submit data to be processed
- User can choose his own data and receives the results after processing
- Not 100% reproducible by user, but enough for comparisons and tests on other data
- Additional advantage: create large test database
Impact

- Papers available online are cited 3 x more often [Lawrence]
- Increased visibility
- More than 100 downloads/month for super-resolution code
- Red-eye removal paper - with online Java code - most popular download in EPFL database
Conclusions

• Reproducible research increases impact
• It helps both yourself and other people
• Start reproducible research now!

More information:
Have a look at
http://lcavwww.epfl.ch/reproducible_research

Or contact me:
Patrick.Vandewalle@epfl.ch