Software Everywhere

- particle physics
- fluid dynamics
- econometrics
- signal processing
- quantum chemistry
- LIDAR archeology
- MRI analysis
- climate & weather
- geophysics
- ...

© CERN, rreis@flickr, rafael grompone, info-nftk@flickr, mohapj@flickr, mario stefanutti, argonne@flickr
Software vs Science?

Research software is not like general-purpose equipment: telescopes, supercomputers, compilers, libraries

Software is made by scientists for scientists, specialized for an experimental process.

Most of the research software is:
- not released
- not published, not reviewed, not cited
- not completely specified
- … and often buggy
Reproducible Research

*Research is reproducible if other researchers can independently obtain the same results from the published material.*

- Theoretical sciences have proofs
- Experimental sciences have procedures
- Computational sciences have ...
  - insufficient descriptions
  - missing parameters
  - missing pre/post processing steps
  - missing data

refs: Claerbout 1992, Donoho 1995, Stodden, Vandewalle
Reproducible Research Initiatives (some)

Journals:
- *Math Programming Computation* requires the code
- *Biostatistics* stamps reproducible articles
- *JMLR* publishes software
- *Geophysics* has some software guidelines
- *Source Code for Biology and Medicine* publishes software
- *Image Processing On Line* focuses on algorithm and software
- *Computing in Science and Engineering* reviews software

Editors:
- *SIAM* updated its supp. material policies to include software
- *ACM* reformed its supp. material copyright policy
- *Elsevier* experiments with “executable papers” and “post-PDF”

Tools and Services:
- *RunMyCode*
- *FLOSShub, mloss/mldata*
- Open data repositories: *DataDryad, Figshare*
Science Code Manifesto

► **Code**: All source code written specifically to process data for a published paper must be available to the reviewers and readers of the paper.

► **Copyright**: The copyright ownership and license of any released source code must be clearly stated.

► **Citation**: Researchers who use or adapt science source code in their research must credit the code’s creators in resulting publications.

► **Credit**: Software contributions must be included in systems of scientific assessment, credit, and recognition.

► **Curation**: Source code must remain available, linked to related materials, for the useful lifetime of the publication.

http://sciencecodemanifesto.org/
OPEN ACCESS

FREE
IMMEDIATE
RESEARCH ARTICLES
ONLINE
AVAILABILITY
RE-USE RIGHTS
Traditional Articles (comput. sciences)

- Publish description of method/algorithm and present results
- The **code** used for obtaining the results is **rarely made public**

Non reproducible results are viewed with skepticism.

Reasons to not distribute the research code?
Traditional Articles (comput. sciences)

- Publish description of method/algorithm and present results
- The **code** used for obtaining the results is *rarely made public*

Non reproducible results are viewed with skepticism.

**Reasons to not distribute the research code?**
- Keep the advantage w.r.t. other researchers
  (better to not publish at all)
Traditional Articles (comput. sciences)

- Publish description of method/algorithm and present results
- The **code** used for obtaining the results is **rarely made public**

Non reproducible results are viewed with skepticism.

**Reasons to not distribute the research code?**
- Keep the advantage w.r.t. other researchers (better to not publish at all)
- Ashamed of the code (no time for cleanup & documentation)
Traditional Articles (comput. sciences)

- Publish description of method/algorithm and present results
- The code used for obtaining the results is rarely made public

Non reproducible results are viewed with skepticism.

Reasons to not distribute the research code?
- Keep the advantage w.r.t. other researchers (better to not publish at all)
  - Ashamed of the code (no time for cleanup & documentation)
  - Prevent incorrect use (by choosing wrong parameters)
Revisit Objectives of Publishing Articles

Picture of a rare utopian community in the act of sharing their research code

vs

Impact Factor

http://www.freeclipartnow.com/
Revisit Objectives of Publishing Articles

Picture of a rare utopian community in the act of sharing their research code

http://www.freeclipartnow.com/

KEY to lure researchers into sharing their code
Revisit Objectives of Publishing Articles

Step 1: Make the code a publication by itself
Revisit Objectives of Publishing Articles

Step 2: Guide community to cite implementations
"IPOL is a research journal of image processing and image analysis. Each article contains a text describing an algorithm and source code, with an online demonstration facility and an archive of online experiments. The text and source code are peer-reviewed and the demonstration is controlled. IPOL follows the Open Access and Reproducible Research models."

http://ipol.im/

For every article, the implementation is:
▶ Reviewed and Published under GPL/BSD license
▶ Following Software Guidelines for correctness, portability, documentation
GOAL: provide a reference implementations of image processing algorithms

- IPOL is not a prototype (running since 2011)
- IPOL is a journal: ISSN; DOI; Int'l editorial committee; ...
- Partnership with a SIAM journal for publication in both journals

- IPOL exists because no other journal did it

- IPOL publishes algorithms, not software. The implementations are here to provide the full details and a way to run the algorithm
IPOL Article Components

- Algorithm description
- Implementation source code

LSD: a Line Segment Detector
Rafael Grompone von Gioi, Jeremie Jakubowicz, Jean-Michel Morel, Gregory Randall

LSD is a line-segment detection algorithm giving subpixel accurate results. It is designed to work on any digital image without parameter tuning. It controls its own number of false detections: on average, one false alarm is allowed per image [3]. The method is based on Burns, Hanson, and Riemann’s method [2], and uses an atrous validation approach according to Desolneux, Moisan, and Morel’s theory [3, 4]. The version described here includes some further improvement over the one described in our original article [1].

Source Code
The ANSI C implementation of LSD version 1.6 is the one which has been peer reviewed and accepted by IPOL. The source code, the code documentation, and the online demo are available at the IPOL web page of this article.

Supplementary Material
Also available at the IPOL web page of this article are two other implementations of LSD, versions 1.0 and 1.1, as well as an example of applying LSD frame by frame, to a video.

Download
- full text manuscript: PDF
- source code: ZIP

Communicated by Lionel Moisan
Demo edited by Rafael Grompone


RIGOUR AND OPENNESS IN 21ST CENTURY SCIENCE — 18
Test the algorithm with new data and explore the parameters without compiling.
IPOL Article Components

- Algorithm description
- Implementation source code
- Web demo interface
- Public archive with original test data
Reproducible+demo → Rigour++

- on-line demo facilitates experimentation
- and leads to a stricter verification of the claims
Reproducible+demo → Rigour++

- on-line demo facilitates experimentation
- and leads to a stricter verification of the claims
IPOL Usage Stats

- 30 articles published with code and demo since 2011
  20 articles in preparation
- 109 citations (cf. Google Scholar)
- 2012: 125000 visits, 13000 code/data downloads
- 2012: 50000 demo runs, 30000 archived runs on original data
IPOL: Good Things

- Reference versions of algorithms
- In-depth analysis of algorithms
- Improvement of the algo. by the reviewer and user tests
- Improvement of the code by the review
- Useful to other researchers

reproducible is good, reusable is better
Challenges

Needs momentum:
- Effort to prepare the code for publication isn't negligible
- Community must learn to cite the implementations

Also: Reproducible ≠ Reusable
Challenges

Needs momentum:
- Effort to prepare the code for publication isn't negligible
- Community must learn to cite the implementations

Also: Reproducible ≠ Reusable

Team-up with: WEB designers & Visualization Scientists?
Follow-up to...

IPOL wouldn't be possible without the support and trust of the authors, reviewers and editors who contributed to it. Lots of help from Paris, Palma, Montevideo, Durham, …

You can be part of it.

http://ipol.im/
edit@ipol.im
discuss@list.ipol.im
@ipol_journal