Software as Science

Computational Science and Software, Open
Access/Data/Source, Reproducible Research, etc.

FOSDEM 2013, FOSS for Scientists
Bruxelles, 2013-02-02

Nicolas Limare
CMLA, ENS Cachan, FR
Image Processing On Line – IPOL
http://www.ipol.im/

Software Everywhere
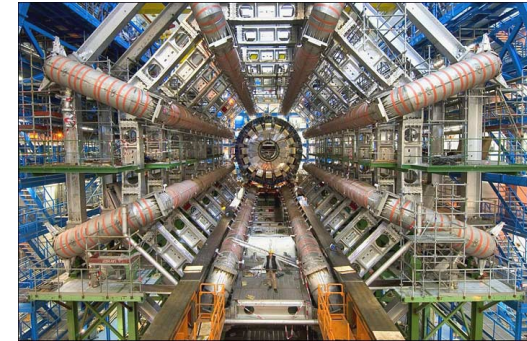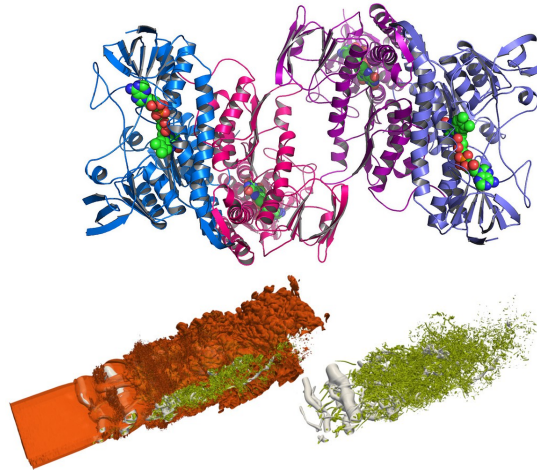
- ▸ particle physics
- ▸ fluid dynamics
- ▸ econometrics
- ▸ signal processing
- ▸ quantum chemistry
- ▸ LIDAR archeology
- ▸ MRI analysis
- ▸ climate & weather
- ▸ geophysics
- ▸ …
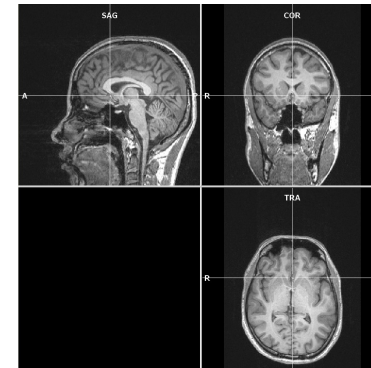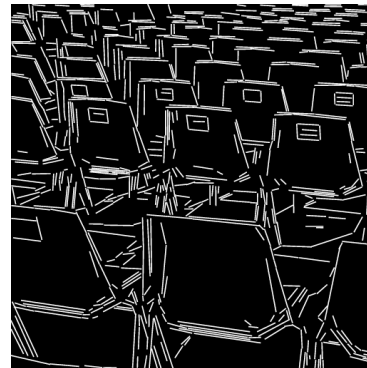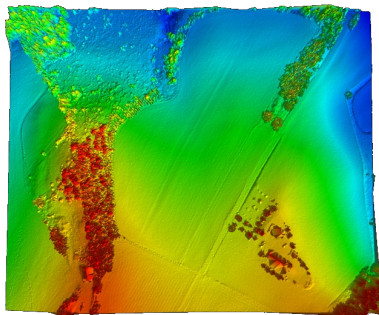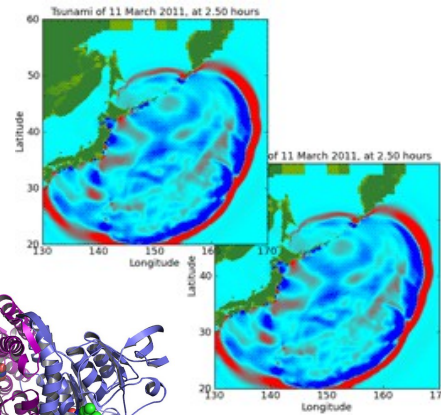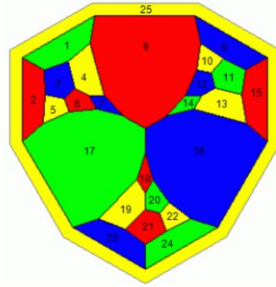
© CERN, rreis@flickr, rafael grompone, info-nftk@flickr, mohapj@flickr, mario stefanutti, argonne@flickr

Software vs Science?

≠ general-purpose equipment: telescopes, supercomputers, compilers, libraries

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

→ compare with science, articles, theorems…

Reproducible (Computational) Research

scientific method

1200 ~ 1800
Roger Bacon, Francis Bacon,
Galileo Galilei, Robert Boyle,
René Descartes, …

→ science needs to be reproduced

reproducible research

1990 ~ …
Jon Claerbout, David Donoho
Serguei Fomel, Randy Leveque,
Davis Bailey, Victoria Stodden,
Juliana Freire, …

→ the science is in the software, data and process

Reproducible Research Initiatives

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"

Data:
- Many institutional data repository: 3TU, Stanford, …
- Open repositories: *DataDryad, Figshare*

Reproducible Research Initiatives

Conferences and Workshops:
- *R4 Workshop*, 2012-04, Orléans, FR
- *eScience 2012 Workshop on Maintainable Software Practices*, 2012-10, Chicago, USA
- *ICERM Workshop on Reproducibility in Computational and Experimental Mathematics*, 2012-12, Providence, USA
- *SINTEF Winter School*, 2013-01, Geilo, NO
- *SIAM CSE13*, 2013-03, Boston, USA
- *Beyond The PDF*, 2013-03, Amsterdam, ND

Tools and Services:
- *RunMyCode*
- *FLOSShub*, *mloss/mldata*, …

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/

IPOL : Image Processing On Line

*"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

IPOL : Web Testing Interfaces

For every article, a web demo interface:
- uses the exact same reviewed code
- runs in real-time (<30s)
- accepts free input and parameters
- saves original data in a public archive

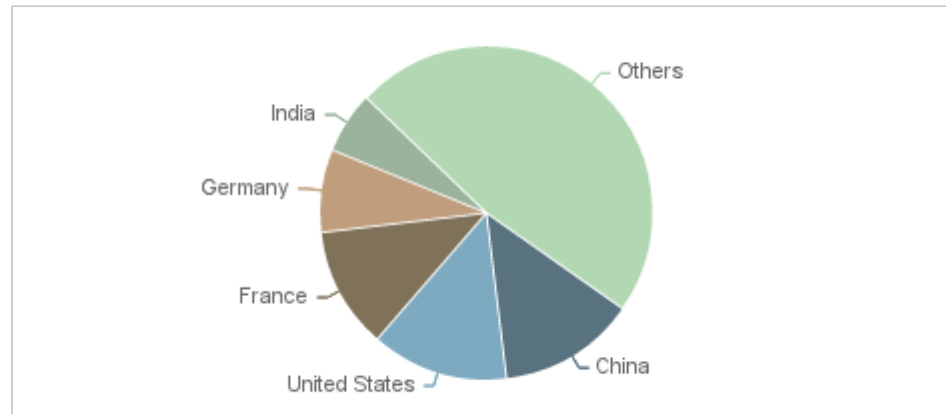→ *"reproducible research as a service"*?
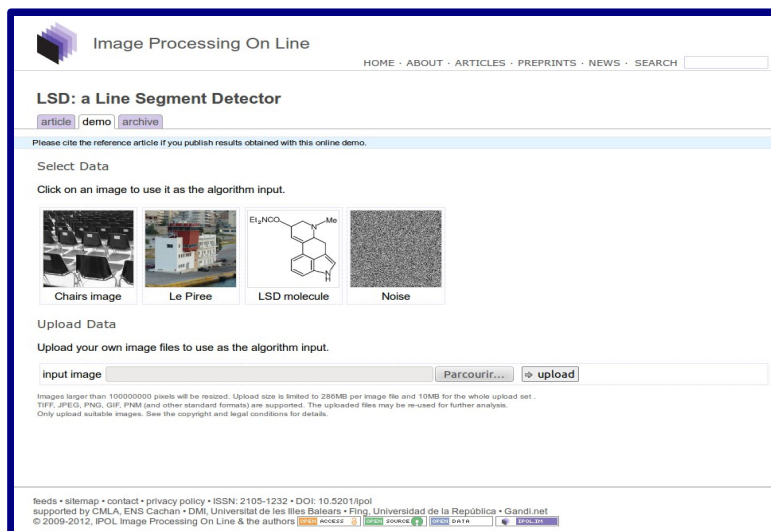
IPOL Usage Stats

▸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

DIY Research Journal with Software

- ▸ISSN, DOI, Title
- ▸Scientific and editorial project
- ▸Editorial & tech committee
- ▸Policy on papers and software:
  what and how to publish, how to review
- ▸Reviewing and publishing tools + backups

- ▸Long term, >3 years
- ▸International collaborations
- ▸"Official" approval
- ▸Some work needed: sysadmin, TeX editing, communication, etc…

- ▸Repository overlay journals? cf. *CCSD/Episciences*

Other Option: Reviewer Pressure

Upgrade needed for the computational research communication standard …
… but publishing policies are very slow to evolve …
… it could be done by you/me/us, the peer-review crowd:

Any reviewer sensible to the need for quality in computational research can right now mention it in their peer-review reports. When asking for minor revisions, include the need for a reproducibility statement, the release of an essential piece of code or data, or more details about the computing environment. This will gradually raise the standards for published papers.

Nicolas Limare, "Upgrading Journal Policies via Peer-Review Influence,"
http://limare.perso.math.cnrs.fr/notes/2013/01/25_policy-reform-via-reviewers/

Other Option: Enforce Good Practices

Reproducibility PI Manifesto by Lorena Barba:

- Teach graduate students about reproducibility
- All research code (and writing) under version control
- Always carry out verification and validation
- Share data, plotting script and figure for main results under CC-BY
- Upload the preprint to arXiv at time of submission
- Release code at time of submission
- Add a "Reproducibility" statement at the end of every paper
- Keep an up-to-date web presence

Reproducibility PI Manifesto. Lorena A. Barba. figshare.
http://dx.doi.org/10.6084/m9.figshare.104539

More Ideas…

- ▸DOI for releases on GitHub &Co.
- ▸Standard citation format for software
- ▸Extending preprint repositories to code and data

- ▸Less NIH implementations, more standard APIs
- ▸Use standard and open languages (MATLAB→Python, cf. SWC)
- ▸Software test tools and services

- ▸Copyright to the authors
- ▸Patent shield for research and experimentation

Follow-up to...

http://ipol.im/
edit@ipol.im
discuss@list.ipol.im
@ipol_journal
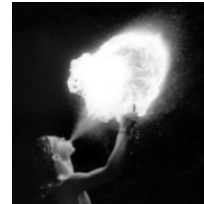
http://nicolas.limare.net/
nicolas@limare.net
@nicolaslimare
→ please talk to me!

and also…
- http://www.runmycode.org/
- http://stodden.net/
- http://reproducibleresearch.net/
- reproducible-research@googlegroups.com
- recherche-reproductible@listes.univ-orleans.fr (FR)
- …