Shining Light into the Black Box

Publication and Citation of Scientific Software Code

Dr. Jens Klump
Deutsches GeoForschungsZentrum GFZ

Imagine mathematics without proofs ...

- Imagine a world in which mathematics papers contain:
  - Lemmas, Theorems, Corollaries
  - No proofs
- Nobody expects to see a proof in a publication, or to ever have to submit one.
- *This is the way it's always been and there are lots of good theorems in the literature, so why change?*

- Suppose somebody started suggesting papers should contain proofs...

LeVeque, 2011

Some objections ...

- The proof is too ugly to show anyone else.
  - It would be too much work to rewrite it neatly so others could read it.
  - It's a one-off proof for this particular theorem, not intended for others to see or use.
  - My time is much better spent proving another result and publishing more papers rather than putting more effort into this one, which I've already proved.

LeVeque, 2011

More objections ...

- I didn’t work out all the details.
  - Some tricky cases I didn’t want to deal with, but the proof works fine for most cases, such as the ones I used in my examples.
  - I discovered some cases actually don’t work, but as long as I don’t mention it nobody will notice.
  - I didn’t actually prove the theorem, my student did.
- And ... the student has since disappeared, along with the proof, but I’m sure it was correct!

LeVeque, 2011

And more objections ...

- The proof is valuable intellectual property.
  - It took years to prove this theorem. Why should I give the proof away freely?
  - The same idea can be used to prove other theorems. I deserve at least 5 more papers before sharing the proof.
  - Someone else might use the ideas in my proof without giving me proper credit.
  - The idea is so great I can commercialize and sell the proof.

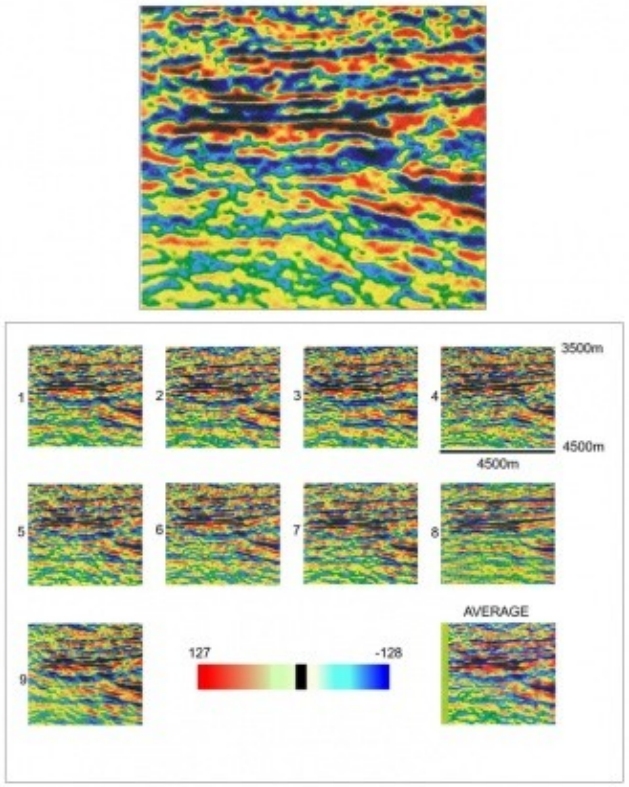
LeVeque, 2011

And even more objections ...

- There are technical difficulties.
  - Including proofs would make papers much longer. Journals wouldn't want to publish them.
  - Referees would never want to have to read proofs. It would be too hard to determine correctness of long proofs and finding referees would become impossible.
  - The proof uses sophisticated mathematical machinery that most readers/referees don't know.
  - My proof uses other theorems with unpublished (proprietary) proofs, so it won't help to publish my proof — readers still will not be able to fully verify correctness.

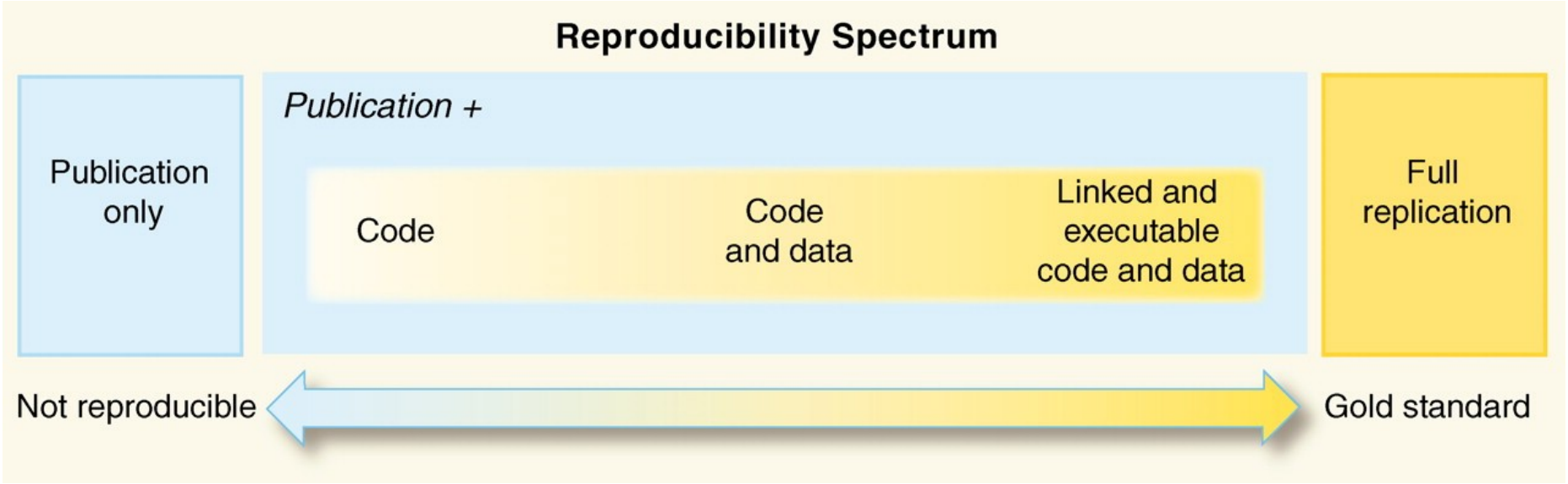
LeVeque, 2011

The Science Challenge



- Example: seismic profiles computed on different hardware- and software stacks.
- Different system configurations produce significantly different results.
- This is in conflict with the basic principle of reproducibility of scientific resu

Is it reproducible?



Peng, *Science*, 2011

Good Scientific Practice

POLICY ACTIONS TO ELIMINATE SOURCE CODE WITHHOLDING IN RESEARCH COMPUTATION	
Institutional support	Publicly funded research institutions and university TTOs must remove organizational impediments to OSS licensing of computer code and embrace a wider variety of methods for exploiting and sharing their intellectual property. Creating a "standard set" of open software licensing tools within and across institutions that includes established OSS licenses would be an important step toward that goal.
Funding policy	Public funding and policy-setting agencies must explicitly and clearly state their strong preference for open dissemination, sharing, and publication of scientist-created software and source code. Although not an absolute requirement in recognition of the enormous diversity of research receiving public funds, the burden of justifying proprietary research products would be left to the applicant.
Publishing requirement	Scientific journal publishers must enact editorial policies requiring, as a condition of publication, that researchers make available new computer source code generated in the course of the research and necessary to reproduce the published research findings. Policies in place at journals already meeting this requirement (*16–18*, *36*) could provide guidance for wider implementation.

Morin, et al., *Science*, 2012

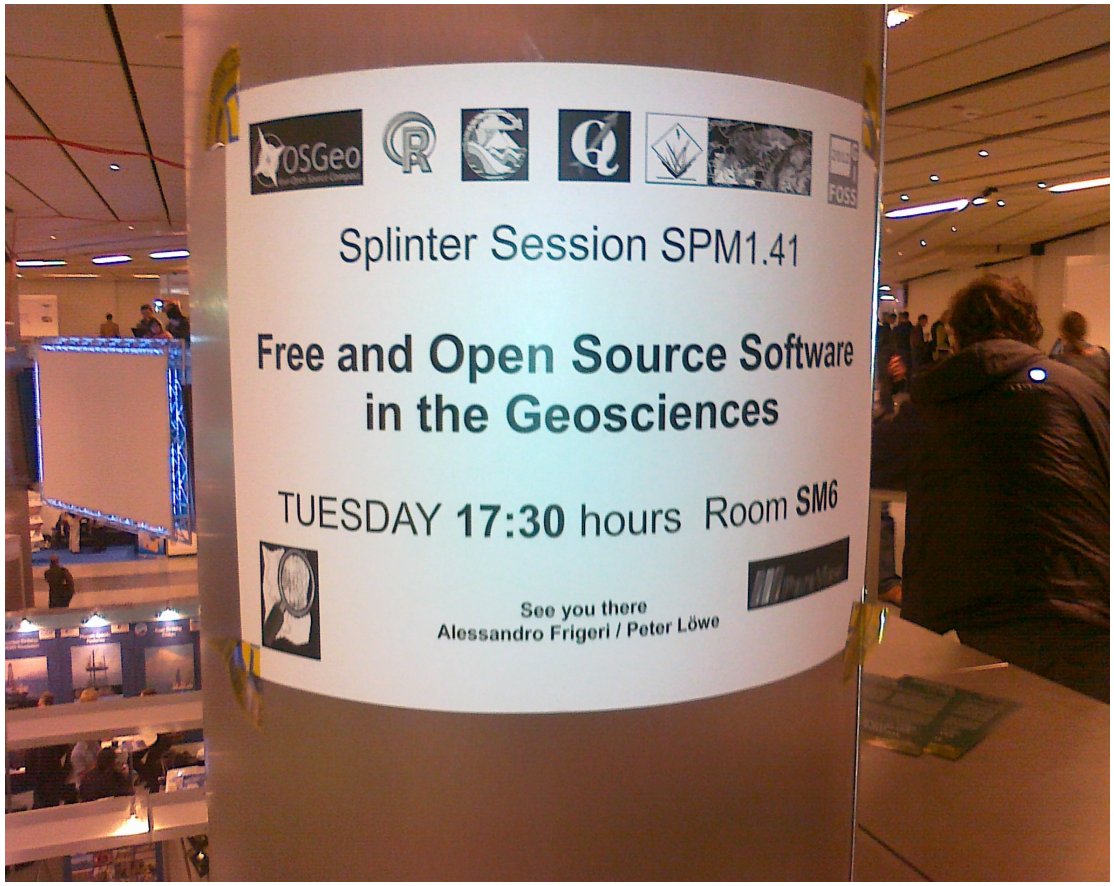
Motivation

In the course of our research create a multitude of algorithms, models, software … in short: code.

Open questions:

- How can I **publish** code?
- How can I **refer** to specific code **versions**?
- Which **licence** models are available and which are suitable in a scientific context?
- Which **distribution** channels and platforms are available for the distribution of source code and documentation of models and software?

Activities



Meetings on FOSS GEO at

- EGU 2012
- GFZ/PIK (2012)
- AGU 2012
- EGU 2013

DFG Proposal

Scientific software distribution platform

- Source code
- Version control
- Packaging
- Persistent Identifiers

Scientific software journal

- Documentation
- Reference/Citation
- Peer-review

Software becomes part of the “record of science”.

Next Steps

- Workshop on software publication with participants from different disciplines (January 2013?)
- DFG proposal for software publication platform and Open Access journal (March 2013)
- Launch of DFG project hopefully later in 2013.

Questions?