"Share and Enjoy": Publishing Useful and Usable Scientific Models

Tom Crick[1], Benjamin A. Hall[2], Samin Ishtiaq[3] and Kenji Takeda[3]

[1] Department of Computing, Cardiff Metropolitan University
[2] MRC Cancer Unit, University of Cambridge
[3] Microsoft Research Cambridge

11 December 2014

Our Computational World

"[Computational techniques] have moved on from assisting scientists in doing science, to transforming both how science is done and what science is done."

Science as an open enterprise, Royal Society (June 2012)

`https://royalsociety.org/policy/projects/science-public-enterprise/`

Ian Holmes
@ianholmes

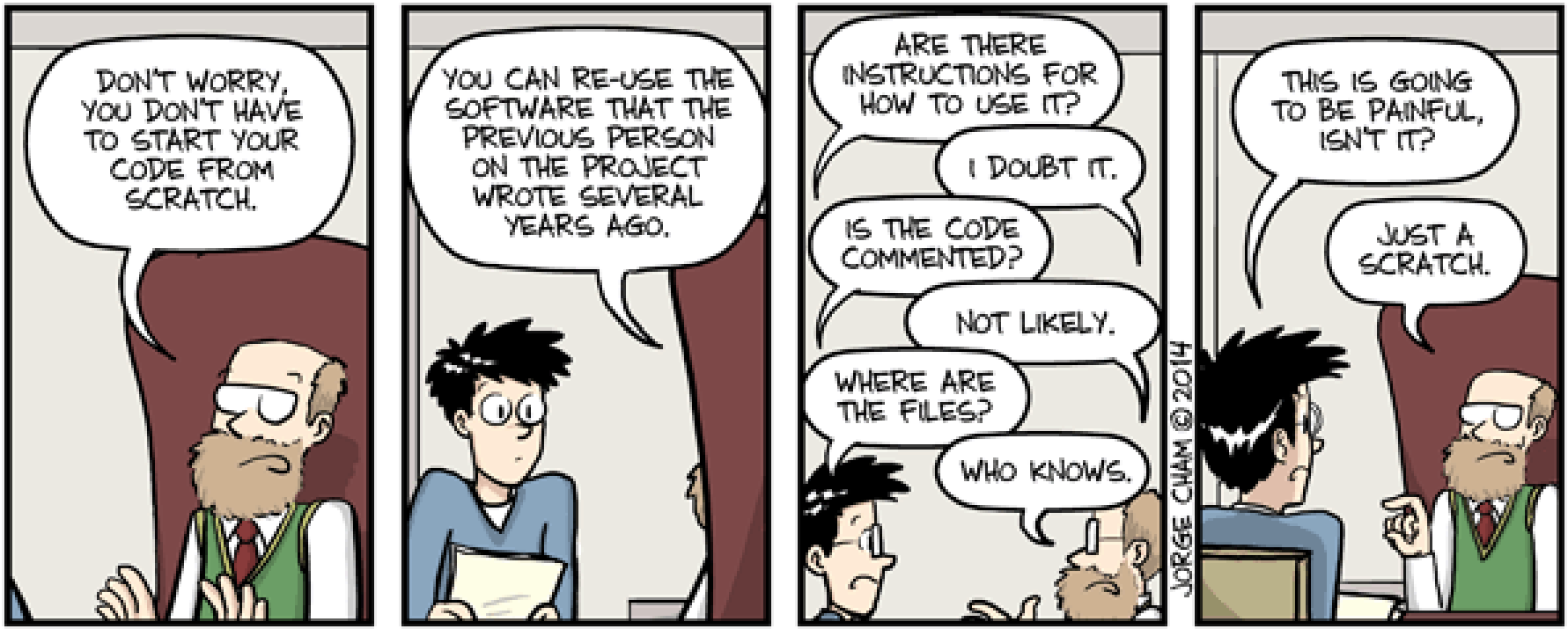
Follow

You can download our code from the URL supplied. Good luck downloading the only postdoc who can get it to run, though #overlyhonestmethods

Reply Retweet Favorite More

RETWEETS 327 FAVORITES 107

4:52 PM - 8 Jan 2013



WWW.PHDCOMICS.COM

Sharing

Two key types of results arise from work done in the computational sciences:

- Models
- Algorithms

Fundamental advantage of computer science and more broadly, computational science: **the unique ability to share the raw outputs of their research as software and datafiles.**

Models, Algorithms and Benchmarks

- Abstraction levels (abstract vs. concrete)
- Benchmark repositories
  (e.g. UCI Machine Learning Repository, Netflix Prize benchmarks, SMT Competition, SV-COMP, Answer Set Programming Competition and the Termination Problem Database.)
- Protocols as scripts (workflow reproducibility e.g. molecular dynamics)
- Performance and scalability (is performance a key issue?)

Models, Algorithms and Benchmarks

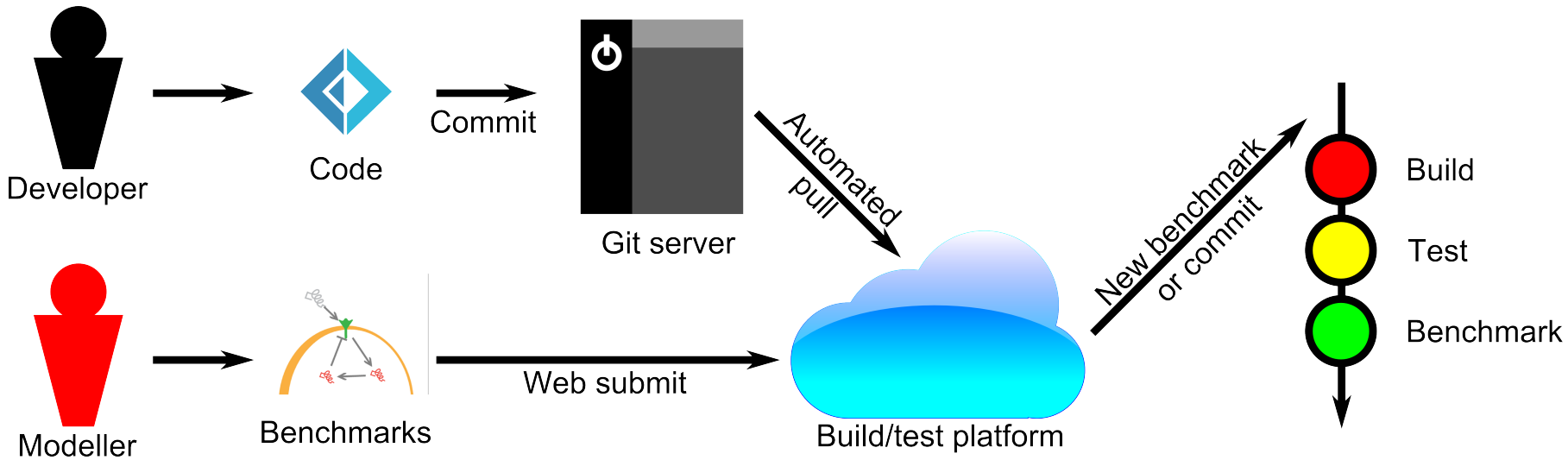
- Abstraction levels (abstract vs. concrete)
- Benchmark repositories
  (e.g. UCI Machine Learning Repository, Netflix Prize benchmarks, SMT Competition, SV-COMP, Answer Set Programming Competition and the Termination Problem Database.)
- Protocols as scripts (workflow reproducibility e.g. molecular dynamics)
- Performance and scalability (is performance a key issue?)

We propose to develop a prototype open software platform which will automate reproducibility for algorithms and models.

A System for Automating Reproducibility in Science

- Linking open software, algorithms and models
- Open and community-curated benchmarks
- Integrated continuous integration system: authoritative source of results for these algorithms running on these benchmarks.

Proposed Workflow



A System for Automating Reproducibility in Science

- Build a cloud service which automatically pulls and compiles code from source repos;
- Run automated tests defined by the developers on the code;
- Perform analysis of benchmark sets supplied by both the developer and external users;
- Provide persistent audit trails for software and benchmarks results;
- Collaborate with key stakeholders in the open software/open data/open access/open science space, as well as key e-infrastructure organisations e.g. GitHub, figshare, SSI, Mozilla Science Lab, Digital Science, etc.
- **Key:** engage with key communities to embed system/workflow and effect cultural change.

Acknowledgements

Software Sustainability Institute

References

- Tom Crick, Benjamin A. Hall, Samin Ishtiaq and Kenji Takeda. *"Share and Enjoy": Publishing Useful and Usable Scientific Models*. In 1st International Workshop on Recomputability, 2014: `http://arxiv.org/abs/1409.0367`
- Tom Crick, Benjamin A. Hall and Samin Ishtiaq. *"Can I Implement Your Algorithm?": A Model for Reproducible Research Software*. In Proceedings of 2nd International Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE2), 2014: `http://arxiv.org/abs/1407.5981`
- Digital Science Catalyst Grant (Nov 2014): `https://github.com/tomcrick/DSCatalyst`
- Microsoft Azure for Research Grant (Dec 2014): `https://github.com/tomcrick/Azure4Research`