Software in Reproducible Research: Advice and Best Practice collected from experiences at the Collaborations Workshop

ACM SIGPLAN TRUST 2014 @ PLDI 2014, 12 June 2014, Edinburgh

Shoaib Sufi, Community Manager, Software Sustainability Institute

shoaib.sufi@software.ac.uk

Supported by EPSRC Pioneering research   *Additional funding* Jisc BBSRC   *Partners* The University of Manchester

The Software Sustainability Institute

A national facility for cultivating world-class research through software

- Better software enables better research

- Software reaches boundaries in its development cycle that prevent improvement, growth and adoption

- Providing the expertise and services needed to negotiate to the next stage

- Developing the policy and tools to support the community developing and using research software

Software Sustainability Institute

"An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures."— *Jonathan Buckheit and David Donoho, paraphrasing Jon Claerbout*

Collaborations Workshop 2014

- March 2014 in Oxford. 66 Attendees.
- Developers, researchers, funders, publishers, managers and project leaders
- Theme – Software in your reproducible research
- Understand what the community see as the problems/solutions
- Unconference, lightning talks, discussions sessions, keynotes, collaborative ideas sessions, hackday

- Collaboration Readiness

- Capability Enhancement

- Advocacy

Collaboration Readiness
Best Practices

- Use Collaborative platforms
  - GitHub
- Share your resources
  - Figshare
- Automate your analysis
  - Make
- Lower the barrier to reproduce your work
  - IPython Notebook
  - Galaxy
- Increase your quality
  - 5* of Research Software - bit.ly/ssi-5-stars
    - Community, Open, Defined, Extensible, Runnable

- ## Ideas for the future
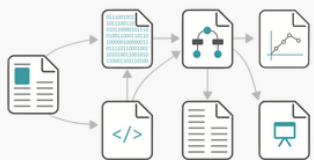 - ### Quality Quality Quality
    - Open Source Health Check
    - Hacker or Slacker
  - ### Connected Research
    - Research Objects; bundle:
      - Code, lab books, software, data, parameters, figures, papers, proposals

    www.researchobject.org

Open Source Health Check
json-editor

json-editor is ready to share :) Check this on ShouldIShareMyCo.de

Well done! This code is licensed under the MIT license.

Great - there's a README file here

The repository doesn't have a CITATION file - adding one could encourage people to reference your code. See here for more information.
The bus factor for this code is approximately 1 over all time
The bus factor in the last year is approximately 1

Weekly commits
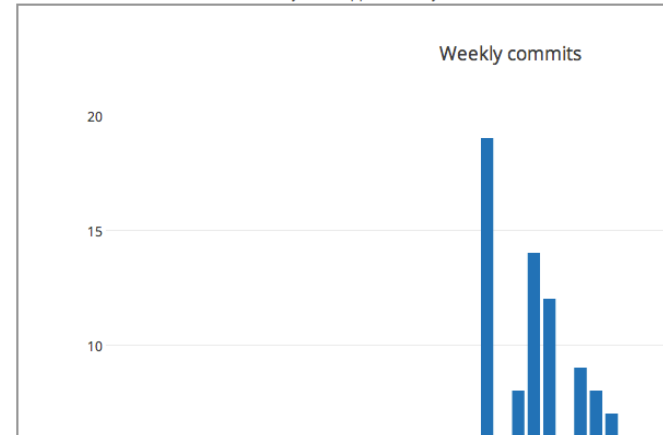
- Auditability of tools, versions and data
    - IPython notebook
    - R Studio
        - Problems in analysis can still happen but they should be <u>easier to identify</u> using this type of tooling
- Packaging reproducible computational experiments
    - Virtual machines (future work: easier interfaces)
    - Recomputation.org
- Choose programming languages with best reproducible features (if possible)
    - Encode the intent of variables (e.g. F#)
    - In built package management (e.g Go)
    - Community Infrastructure (libraries, support etc.)

- Easier for the non-expert
  - Visual version of Make
  - Hiding version control systems behind Apps/Gui's – e.g. genome annotation backed up by Git
- Digitisation of existing resources
  - Chemical structures identification from lab books
- Spreadsheets as web accessible datasources
  - Easy to publish, programmatically accessible (e.g. webservices)

Advocacy
Best Practice

- Promoting and employing relevant Training
  - Core computational competencies
    - Early career & PhD – www.softwarecarpentry.org - bootcamps
  - Reproducible Research training
    - Ongoing challenge to ingrain in research culture
      - Learning task get others to reproduce others work
- How to fund a software developer
  - Making the case in 'Pathways to Impact'
  - Check other successful grants that have done this
  - Suggest software development savvy reviewers if possible
- Support relevant Policy drives
  - Research Software Engineers
    - Establishing a term and a community – www.rse.ac.uk

Software Sustainability Institute

Advocacy
Future Ideas

- Reproduciliteracy
  - Ambassadors and material to promote reproducible research to particular domains
    - www.software.ac.uk/fellows
- Open Science News
  - Aggregating blogs, news etc for those interested in reproducible computation research
- Citing Software
  - Standards and tooling needed
- Publishers stick
  - Mandating the publishing of code and data
- Tracking Software development time
  - So that time taken better understood by funders and PI and this more justifiable
- Impact of software on research
  - ScienceToolBox sciencetoolbox.org
    - Linking developers with software with citations

Recap

- Key Reproducible Research Themes from CW14:
    - Collaboration readiness
    - Capability enhancement
    - Advocacy
- Best Practice – things that can be done today !
- Future Ideas – things worth/being trying/tried.

Take Home

- Advocacy – only through ongoing education and promotion will we get the **culture change** need to make reproducible research practices ubiquitous ...

- Yes – reproducibility does add overhead – but even some reproducible practices are better than none – it does not have to be perfect to be a huge improvement ... it's about increasing the **quality, trust and value** of and in research outcomes.

- Plus – sooner or later Journals/Research will mandate this – it's worth being ahead of the curve :)

Software Sustainability Institute
supporting software in research
-- powered by you ! --

www.software.ac.uk

www.software.ac.uk

- **Get involved**

Read, re-tweet(*@SoftwareSaved*), ask, blog, news, Fellows, funding, software carpentry, policy, read/write resources

- **Upcoming**
  - Research Software Engineers workshop – September 2014
  - IP and Licensing workshop – December 2014

- Any questions
    - ?