Research Software Engineering at Sheffield

SOFTWARE IS ESSENTIAL FOR RESEARCH
Do you use research software?
(382 responses)

91% Yes
9% No
Do you develop research software?
(382 responses)

- Yes: 27%
- No: 73%
Do you feel you have received sufficient training to develop reliable software? (103 responses)

- Yes: 31%
- No: 69%
A Scientist's Nightmare: Software Problem Leads to Five Retractions

Greg Miller
See all authors and affiliations

Science 22 Dec 2006
Vol. 314, Issue 5807, pp. 1856-1857
DOI: 10.1126/science.314.5807.1856

Until recently, Geoffrey Chang's career was on a trajectory most young scientists only dream about. In 1999, at the age of 28, the protein crystallographer landed a faculty position at the prestigious Scripps Research Institute in San Diego, California. The next year, in a ceremony at the White House, Chang received a Presidential Early Career Award for Scientists and Engineers, the country's highest honor for young researchers. His lab generated a stream of high-profile papers detailing the molecular structures of important proteins embedded in cell membranes.

Then the dream turned into a nightmare. In September, Swiss researchers published a paper in Nature that cast serious doubt on a protein structure Chang's group had described in a 2001 Science paper. When he investigated, Chang was horrified to discover that a homemade data-analysis program had flipped two columns of data, inverting the electron...
We need research software to be:

- **Correct** tested, testable
- **Understandable** transparency, sustainability
- **Reproducible** research methods & ethics
- **(Re)Usable** (unlocking knowledge)
- **Easy to change** (new knowledge!)

**Why is this so hard?**
The “Research Software Engineer”
What do RSEs do?

- Build **quality** software - as **part of** research projects

- Help researchers:
  - Learn computing skills
  - Establish good practices
  - Navigate computing and data infrastructure

Photo by Goutam1962, under CC-BY
Why RSE Sheffield?

For RSEs

➔ Stable careers
➔ Peer support
➔ Recognition & reward
➔ Professional development

For Research Projects

➔ Fractional support
➔ Flexible access to expertise
➔ Access to niche skills
➔ Skill sharing between projects

For Researchers

➔ Help, support and mentoring
➔ Formal training
➔ Reproducible process and best practice
RSE Sheffield

https://rse.shef.ac.uk/contact/team/
Some projects we’ve been involved with

1. **Accelerating software** using Graphical Processing Units (GPUs)
   - Rob Chisholm
2. **Epidemiological modelling software**
   - Bob Turner
3. **Machine Learning and AI**
   - Twin Karamkharm
4. **Impact from software**
   - Will Furnass (in place of David Wilby)
Improving Performance using GPUs
For tumor modelling
What were the impacts?
April 2020: How could we help?

Rapid Assistance in Modelling the Pandemic: RAMP

Fellows of the Royal Society are using their expertise to track the spread of Coronavirus COVID-19.

SCRC docs

Data
Available data
Data Sources and Products
Model Parameters

SCRC data pipeline
Data Registry
Julia
R
Python

Research

Who are we?

The Scottish COVID-19 Response Consortium is formed of dozens of individuals from over 30 academic and commercial organisations.

Researchers in these organisations jointly responded to a call by the Royal Society to develop more epidemiological models of COVID-19 spread - RAPID ASSISTANCE IN MODELLING THE PANDEMIC: RAMP - in order to develop a more robust and clearer understanding of the impacts of different exit strategies from lockdown. Scientists from several other organisations across the UK and abroad have now joined the consortium to provide additional expertise in specific areas.
Neil Ferguson's Imperial model could be the most devastating software mistake of all time

The boss of a top software firm asks why the Government failed to get a second opinion from a computer scientist

DAVID RICHARDS AND KONSTANTIN BOUDNIK
31 May 2020 - 12:22am

Critiqued coronavirus simulation gets thumbs up from code-checking efforts

Influential model judged reproducible – although software engineers called its code ‘horrible’ and ‘a buggy mess’.

Dharmesh Singh Chauva
What did RSE do?

https://guides.github.com/introduction/flow/
What were the impacts?

SCRC Software checklist

This checklist has been developed by the Scottish Covid-19 Response Consortium and has been used to assess the software engineering aspects of "model readiness" for our six epidemiological models. This is intended to be part of a broader scheme to evaluate and describe suitability of model results to be used in decision making so this checklist covers only the software implementation and assumes that other documents cover questions about model validation, quality of science, data provenance and quality and policy readiness.

Software Details

Model / software name

7 pieces of modelling software; rapid reproducible data pipeline; follow on grant
Machine Learning (ML) and Deep Learning (DL)

- Statistical **models** that learns from **data**.
- DL is a sub-category of Machine Learning that uses ‘deep’ Neural Networks.
- Has huge impacts across most fields in industry and academia.
ML & DL: What’s involved and how can RSEs help?

- Algorithms and Workflow
- Data Collection, Curation and Pipelining
- Computation Resource

Training, consultancy, development

Hardware consultancy, HPC support & on-boarding
ML & DL: Training

● Introduction to Deep Learning
  ○ 1-day workshop with Google Colab
  ○ Practical labs using Tensorflow Keras
  ○ Available in Python and R
ML & DL: Projects

- Rumour Veracity - DL-based service for classifying tweet misinformation and stance
  - Development and deployment of service
  - Twitter rumour veracity classification model implemented in PyTorch
  - Model-retraining infrastructure
ML & DL: HPC Support

- Computing resource consultation
- HPC
  - User onboarding
  - Software support
  - Technical consultation
  - Associated systems:
    - Sheffield
      - ShARC
      - Bessemer
    - Tier-2
      - JADE
      - JADE II
      - Bede
Research Software Impact Enhancement

E.g.

- Packaging/containerising
- Streamlining code - readability, collaboration, usage
- Web apps & APIs
- Measuring software usage & impact
- and more!
Case Study: Impact Enhancement

ML-based classifier → Elasticsearch Instance → Database → Python web app

Social Media Analytics Services

https://firstdraftnews.org/long-form-article/data-deficits/
Partnerships and sustainability

Shorter-term (days); Responsive support

Longer-term (months/years); Embedded collaboration in projects

ITS R&I Support team

Research Software Engineering team

RSE team not an island: work with ITS R&I, Library, Bioinformatics Core, etc.
What we can offer

● Collaboration (Projects)
  ○ RSEs working as part of grant-funded projects (%FTE or per day)
  ○ Pre-award consultation required
  ○ Some support for post award projects

● Education (collaborative support co-run with ITS R&I)
  ○ Training, mentoring and training development
  ○ Free Code Clinics: consultation service

● Community (for RSEs and Research Developers)
  ○ LunchBytes: monthly series of talks from the research community on research software, data and infrastructure.
  ○ Collaborative coding events e.g. ReproHacks, GPU Hackathon, Hacktoberfest
  ○ Mailinglist, Newsletter and blog
Getting in touch

rse@sheffield.ac.uk | @RSE_Sheffield | https://rse.shef.ac.uk

Q&A