Virtual SEFARI HPC and Data Sharing Workshop: March 23-25th 2021

For details on the practical sessions, please see Page 2.

MARCH 23rd: Day 1
13:00 General introduction
13:10 Practical session
13:55 *Break*
14:00 Practical session
14:55 *Break*
15:00 Practical session
15:45 *Break*
16:00 Kristina Riemer (University of Arizona): Talk: "Practical advice for levelling up open science skills", followed by discussion
17:00 End

MARCH 24th: Day 2
13:00 Practical session
13:55 *Break*
14:00 Practical session
14:45 *Break*
15:00 Edward Wallace (University of Edinburgh): Talk: "Fundamental skills for open science", followed by discussion
16:00 End

MARCH 25th: Day 3
13:00: Reproducibility (e.g. pipelining, workflow engines)
13:55 Break
14:00: Version Control: GitHub and GitHub case studies
14:30: Catch-up time and question and answer.
14:55 Break
15:00: Curating and sharing data on an HPC: Intro & FAIR data principles and 15min Discussion: Data sharing and collaborations: adding value
15:45: Break
16:00 Data Sharing & Open Science" Talk: FAIR working in practice, for researchers, PI's and institutions: Prof Andrew Millar (FRS FRSE) The University of Edinburgh
(Note this talk will be given using a separate Webex link)
17:00: End

Material for discussions: Data sharing war stories discussions – what works; what doesn't

Practical Sessions

Practical sessions will be split up across the days (and intermixed with the formal talks/discussion sessions). They will cover:

- An Introduction to Linux
    - connecting to the cluster
    - running basic commands in a non-graphical environment
    - navigating a Linux filesystem
    - viewing and editing files
- Shared Analysis Environments
    - HPC architecture overview
    - basic shell scripts
    - job scheduling
    - job types (single, multi-core, parallel, etc)
- Resource Allocation / Best Practice
    - storage tiers
    - monitoring jobs
    - allocating cpus/memory
    - cluster etiquette
- Reproducible Research
    - what *is* reproducible research?
    - working with conda package management
    - running and building Singularity containers

Guest Speaker Information

Dr. Kristina Riemer is a scientific programmer in the Data Science Institute at the University of Arizona. She builds and documents software tools for scientific research, provides computational training for other researchers, and runs crop ecosystem simulation models, and generally promotes reproducible and open approaches to improve the accuracy and speed of scientific progress. She completed her PhD in ecology at the University of Florida, working with Dr. Ethan White.

Dr Edward Wallace is a Sir Henry Dale Fellow (Group Leader) in the Institute of Cell Biology at the University of Edinburgh. His Fellowship, started in 2018, is supported by The Wellcome Trust and The Royal Society. His research focusses on using genome-scale datasets to understand how fungi dynamically reorganize their RNA and protein to adapt to environmental change. Alongside his research, Dr. Wallace is an open science advocate and teaches data literacy to scientists, working with Edinburgh Carpentries.

Professor Andrew Millar originally studied the 24-hour, biological clocks in plants and algae for 30 years. His group's commitment to Open research data management is retained as the 'BioRDM team' at the University of Edinburgh. Andrew has been seconded to the Scottish Government since October 2018 as the Chief Scientific Adviser for Environment, Natural Resources and Agriculture. He has also served on the Council of UKRI-BBSRC since April 2018, where his work has included a review of data-intensive bioscience.