Reproducible research using electronic notebooks

Jake Surman[1]

David So Ri Jung[2]

[1] University of NSW, Sydney, Australia, jake@unsw.edu.au
[2] University of NSW, Sydney, Australia, david.jung@unsw.edu.au

The ability to reproduce a finding is an essential part of good research. However, reproducibility is lost when the backstory that outlines how the data was generated is either not locatable, indecipherable, lost, erroneous, or unrecorded. Electronic lab notebooks (eNotebooks) are great tools to enable reproducible research with benefits including data protection, ease of contextualisation and consolidation of digital data, and searchability. However, eNotebooks still rely on the proactiveness and competency of the researchers to record the right information at the right time, in the right place.

In this presentation, we will explore automated workflows in eNotebooks to further enable reproducible research by reducing the burden on researchers to record procedural data and metadata. As a pilot approach, we investigated the use of application programming interfaces (APIs) to upload high performance computing (HPC) jobs and standard outputs to an eNotebook in LabArchives. This approach sought to minimise user input while automating the capture of information on how and when jobs were run on the HPC systems. The same approach can also be applied to Git repositories, whereby specific versions of software used for a research activity could be pushed through and stored in the eNotebook alongside the generated data.

Applying a similar approach to physical science research and in collaboration with the Mark Wainwright Analytical Centre at UNSW Sydney, a major research instrumentation facility, we explored scripts which saved instrument metadata onto an eNotebook when a researcher collects data using a centrally managed instrument.

By recording greater details of the research activities, either automatically or with minimal effort, these workflows increase the quality and reproducibility of the research and establish the eNotebook as a single source of truth for all research procedural data and metadata.