Challenges and solutions in replicability and provenance tracking for simulation projects

Andrew P. Davison
Unité de Neurosciences, Information et Complexité
Centre National de la Recherche Scientifique
Gif sur Yvette, France
andrew.davison@unic.cnrs-gif.fr
http://www.andrewdavison.info
@apdavison

Background
Provenance of computational simulations is a challenging topic. It is essential to ensure that the results obtained from these simulations are reproducible. However, the lack of tools to capture and store the provenance information makes it difficult to reproduce the results. The aim of this project is to develop a tool to capture and store the provenance information of simulation runs.

The tool should be able to capture the following information:
- The code used to run the simulation
- The input files used
- The parameters used
- The output files generated
- The system configuration used
- The time taken to run the simulation

Requirements
- Support for different simulation frameworks
- Support for different programming languages
- Support for different platforms
- Support for different data formats
- Support for different output formats

Why did I do that?
I thought I used the same parameters but I'm getting different results
I can't remember which version of the code I used to generate figure 6
The new student wants to reuse that model I published three years ago but he can't reproduce the figures
It worked yesterday
Why did I do that?