The Practice of Reproducibility: How computational reproducibility emerges from researcher workflow.

Ana Nelson

Extended Abstract:

This will be a practical talk showing a reproducible process for generating a mock research paper based on an agent-based simulation using open source tools like Docker (docker.io) and Dexy (dexy.it).

Agent-Based Modelling researchers have more to gain than most programmers by employing good reproducible development and documentation practices. Not only do ABM researchers have to maintain simulation code, they have to analyze generated data and write up results, with the possibility of having to completely redo this analysis if they update their models and generate new data. A reproducible workflow based on open source tools not only benefits the individual researcher or team using it, it also benefits the whole community, since anyone else can easily reproduce results and re-use code.

This talk will describe one example of a fully automated (and therefore reproducible) workflow, including downloading and installing all necessary software and simulation source code, running the simulation to generate data, running analysis scripts to generate plots and calculated data, and embedding these into documents. We will see how changing simulation code and re-running the workflow results in automatically updated documents: plots will change to reflect the newly-generated data, and software documentation will show the updated source code.

There are two goals for this talk. First, to make researchers aware of the possibilities of a fully automated and reproducible workflow. Second, to present one possible suite of open source tools for implementing such a workflow.

Dexy (dexy.it) is a tool for automating projects. It's similar in spirit to GNU make, but with lots of document-related features. Dexy can automate running scripts and processes, and it also facilitates the embedding within documents of generated artifacts like source code, data and plots. It's easy to add Dexy to an existing project, and Dexy supports multiple document formats and programming languages. It's a software documentation tool and a reproducible research tool, and it was directly inspired by the challenges of Agent-Based Modelling research.
Docker (docker.io) is a convenient tool for creating isolated lightweight "containers" on Linux. By creating a separate container for each project you can keep your primary operating system clean of clutter and also create a reproducible workspace for each project. Dockerfiles, used to configure the setup script for each container, are easily readable by non-Docker users, so they can act as either an executable script, or as a testable list of software dependencies.

**Speaker Info:**

Ana Nelson works as a software consultant in the San Francisco Bay Area. Previously, she completed a Ph.D. in economics at Trinity College, Dublin and attended Swarmfest as a graduate student. She is the author of the open source Dexy package for project and document automation.