Overview

Scripting allows shell users to automate repetitive tasks, thus improving reproducibility, reducing the risk of errors, freeing time, and avoiding boredom.

Bioinformatics analysis pipelines may involve dozens or hundreds of steps that are each carried out by different command-line programs: assembling these into scripts allows users to treat whole pipelines as if they were ordinary shell commands.

This course targets users who have basic knowledge of interactive shell use (such as taught in the SIB's First Steps with UNIX in Life Sciences and are interested in moving from interactive to automated tasks.

At the end of the course you should understand:

* the main syntactic constructs of Bash (tests, conditionals, loops, functions)
* how to read input and write output
* how Bash stores and processes data (including the various kinds of expansion)
* how to pass and parse command-line arguments and options, as well
* how to to assemble individual analysis steps into reproducible, automated pipelines

Technical

A laptop with a command line terminal and a recent version of Bash (4.0 or newer should be ok) installed. The alternative for Windows users is to download a virtual machine image and thus at least 10 Go free on hard disk and 4 Go of RAM are required.

Prerequisites

Knowledge / Skills

Participants are expected to know how to use a Unix shell interactively, i.e., moving around the filesystem, understanding pathnames, lauching programs that work on data, redirecting output to files or other programs, etc. They are also expected to have some familiarity with basic shell utilities like grep, cut, tr, etc (although no expert skills are required). Notions of programming (in any language) will be useful, but not essential. Finally, some proficiency with a text editor is a must.