Linux command line and nf-core introduction

Edoardo Piombo

Umeå Plant Science Centre

A collaboration between

UMEÅ UNIVERSITY

SLU SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES

There are different ways of interacting with computers

Graphical user interface (GUI)
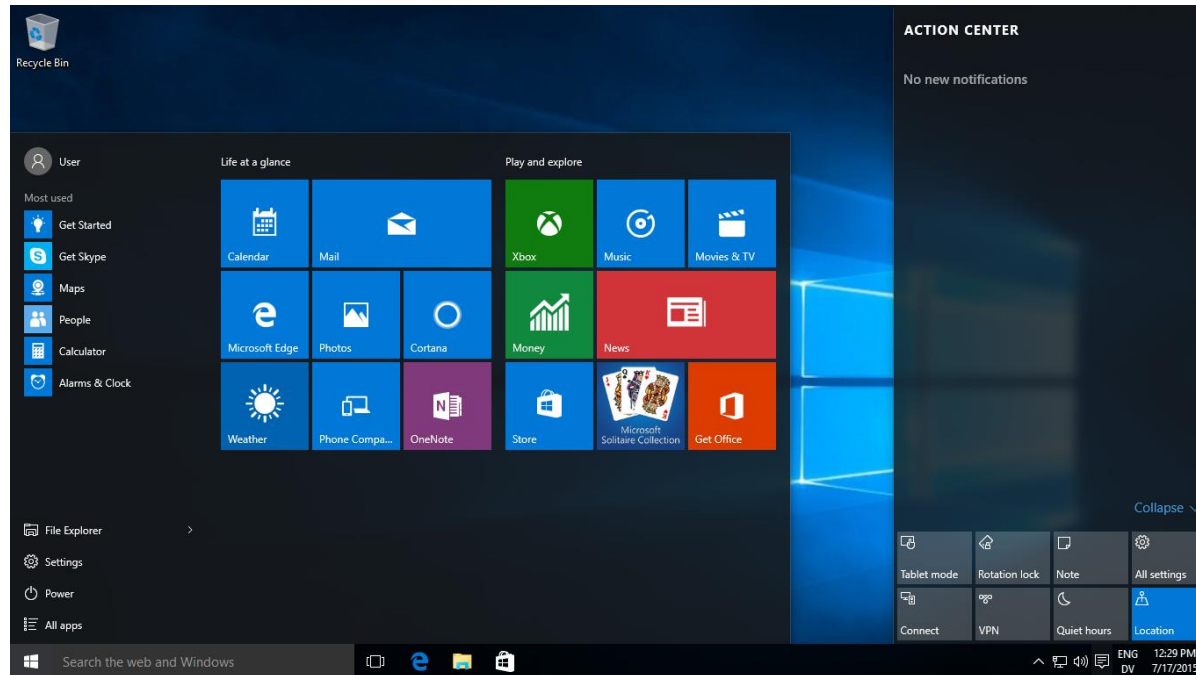
Command-line (or text) interface (CLI). Also called **shell**

Umeå Plant Science Centre

Linux

- Operating system (OS)

- Microsoft Windows, MacOS, and **Linux**

- Linux – Versatile, open source, free,
  used on supercomputers.

- The most common way to run complex bioinformatic
  analyses, on supercomputers or your own pc,
  is to use the **Bash shell.**

User

Application

Operating system

Hardware

Umeå Plant
Science Centre
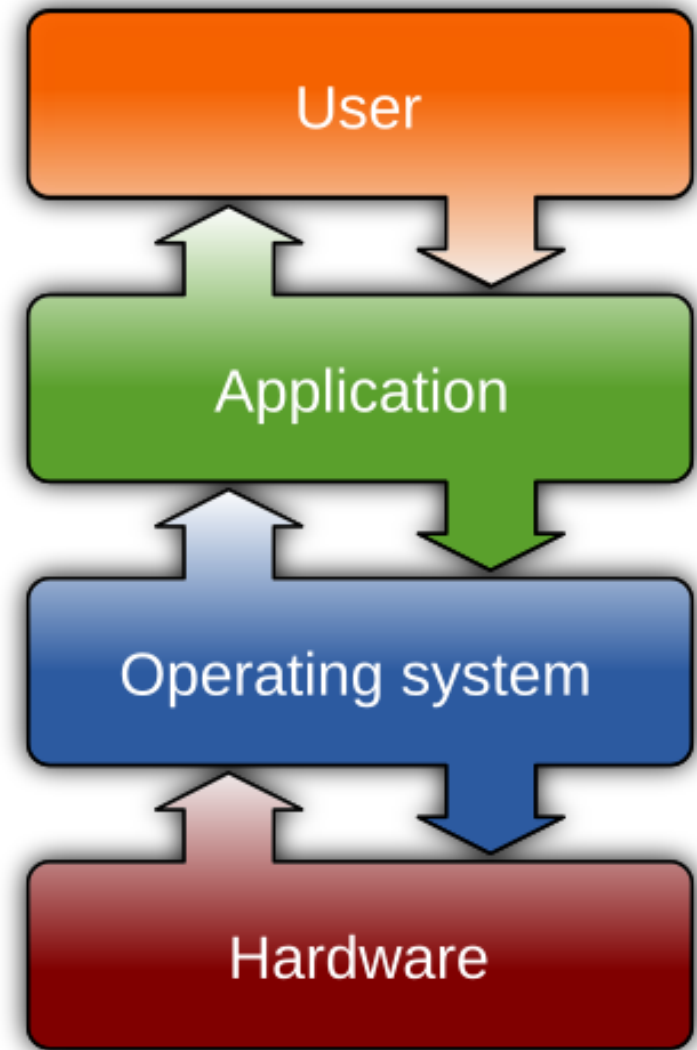
What we will do today

- Tutorial about the bash shell

Break

- Learning about **nf-core**

- Exercise about **nf-core**

Umeå Plant
Science Centre

Downloading this presentation

- Go to this link:
  https://github.com/EdoardoPiombo/nf-core-introduction

- Click on "Presentation.pdf"

  or "Presentation.pptx"

EdoardoPiombo Uploaded presentation		84b9ce6 · 27 minutes ago	9 Commits
📁 raw	Delete raw/fastq_1		31 minutes ago
📁 solutions	Add files via upload		2 weeks ago
📄 LICENSE	Initial commit		2 weeks ago
📄 Presentation.pdf	Uploaded presentation		27 minutes ago
📄 Presentation.pptx	Uploaded presentation		27 minutes ago
📄 README.md	Update README.md		2 weeks ago
📄 example_sample_sheet.csv	Added empty files to test pipelines		2 weeks ago
📄 genes_coordinates.gff	Added empty files to test pipelines		2 weeks ago
📄 genome_file.fasta	Added empty files to test pipelines		2 weeks ago
📄 install_nextflow.sh	Create install_nextflow.sh		2 weeks ago

Umeå Plant
Science Centre

Downloading this presentation

Click on "download raw file" and then open the downloaded presentation.

Umeå Plant
Science Centre

Let's open a terminal

Windows people should do the following:

- download moba X term from here: https://mobaxterm.mobatek.net/
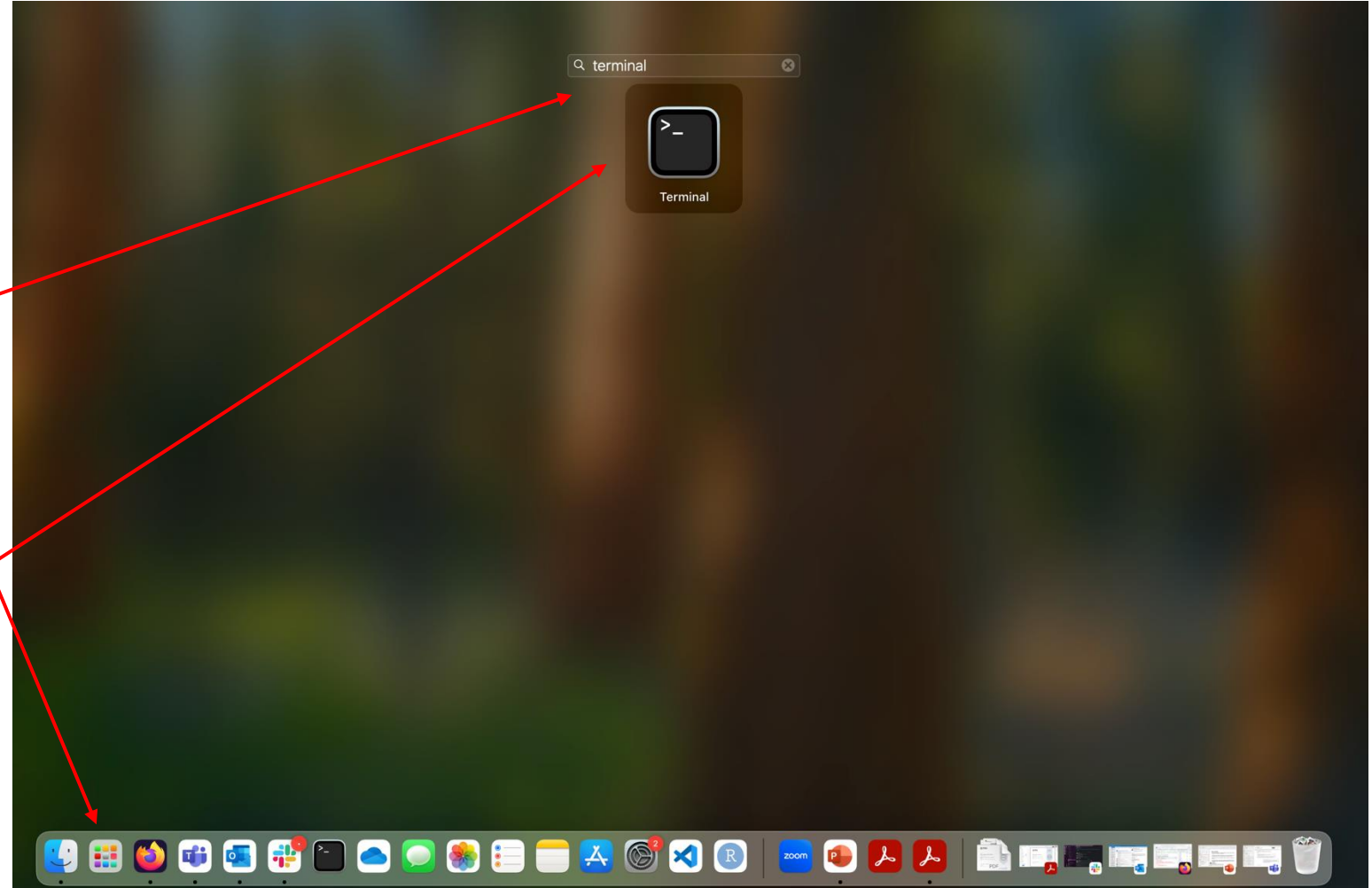
- Follow the instructions to make a bash session.

People using Macs or Linux just need to open a terminal:

- Mac: https://support.apple.com/guide/terminal/open-or-quit-terminal-apd5265185d-f365-44cb-8b09-71a064a42125/mac

- Linux: https://ubuntu.com/tutorials/command-line-for-beginners#3-opening-a-terminal

Umeå Plant Science Centre

Open a terminal: Mac

- Open launchpad

- Write "terminal" in the search bar.

- Click on Terminal

Umeå Plant Science Centre

Open a terminal: Mac

- You should see something like this

```
Last login: Wed May 21 08:53:53 on console
(base) edoardo@client227-209 ~ %
```

edoardo — -zsh — 80×24
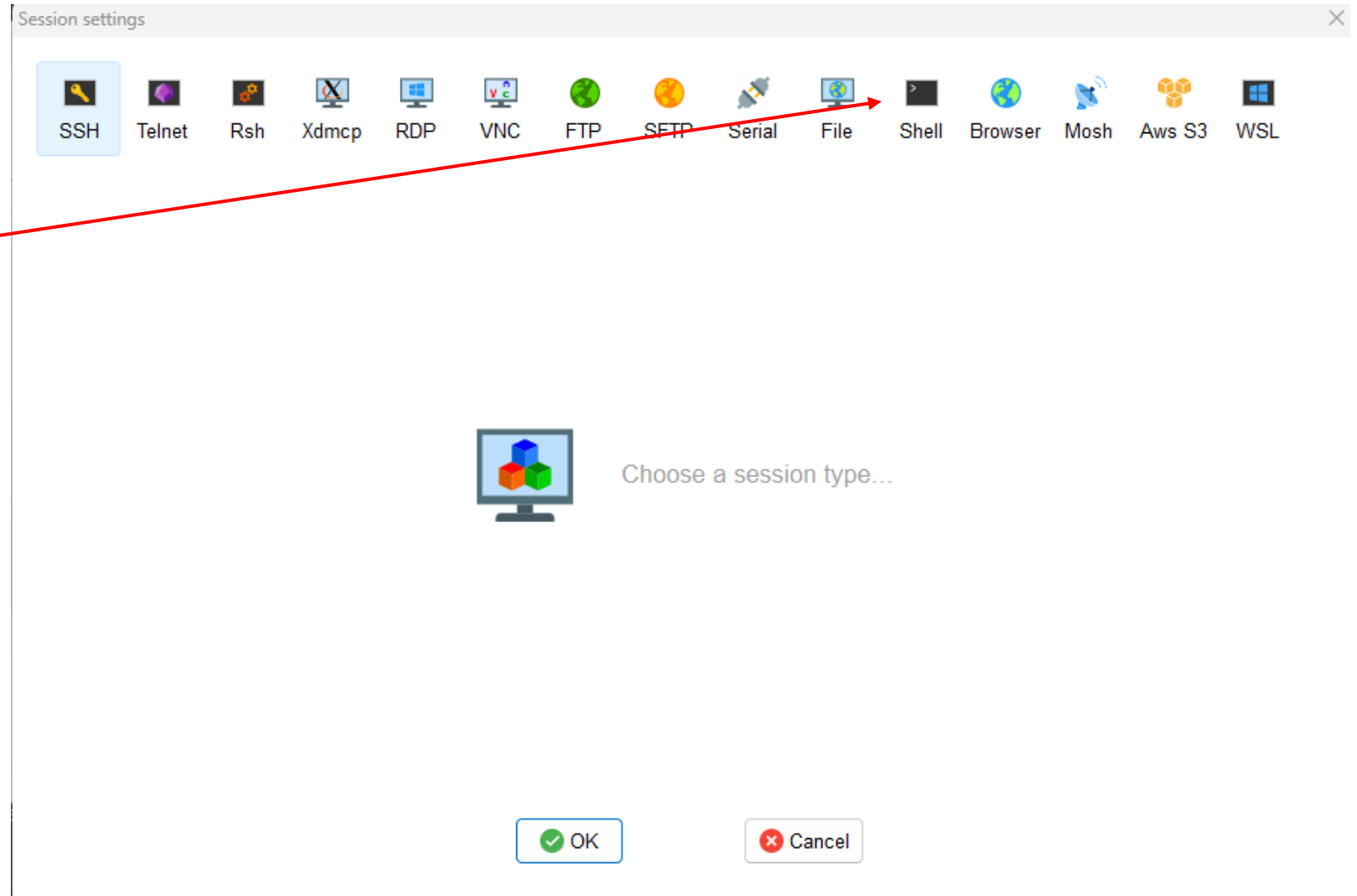
Umeå Plant
Science Centre

Open a terminal: Windows

- Open MobaXterm
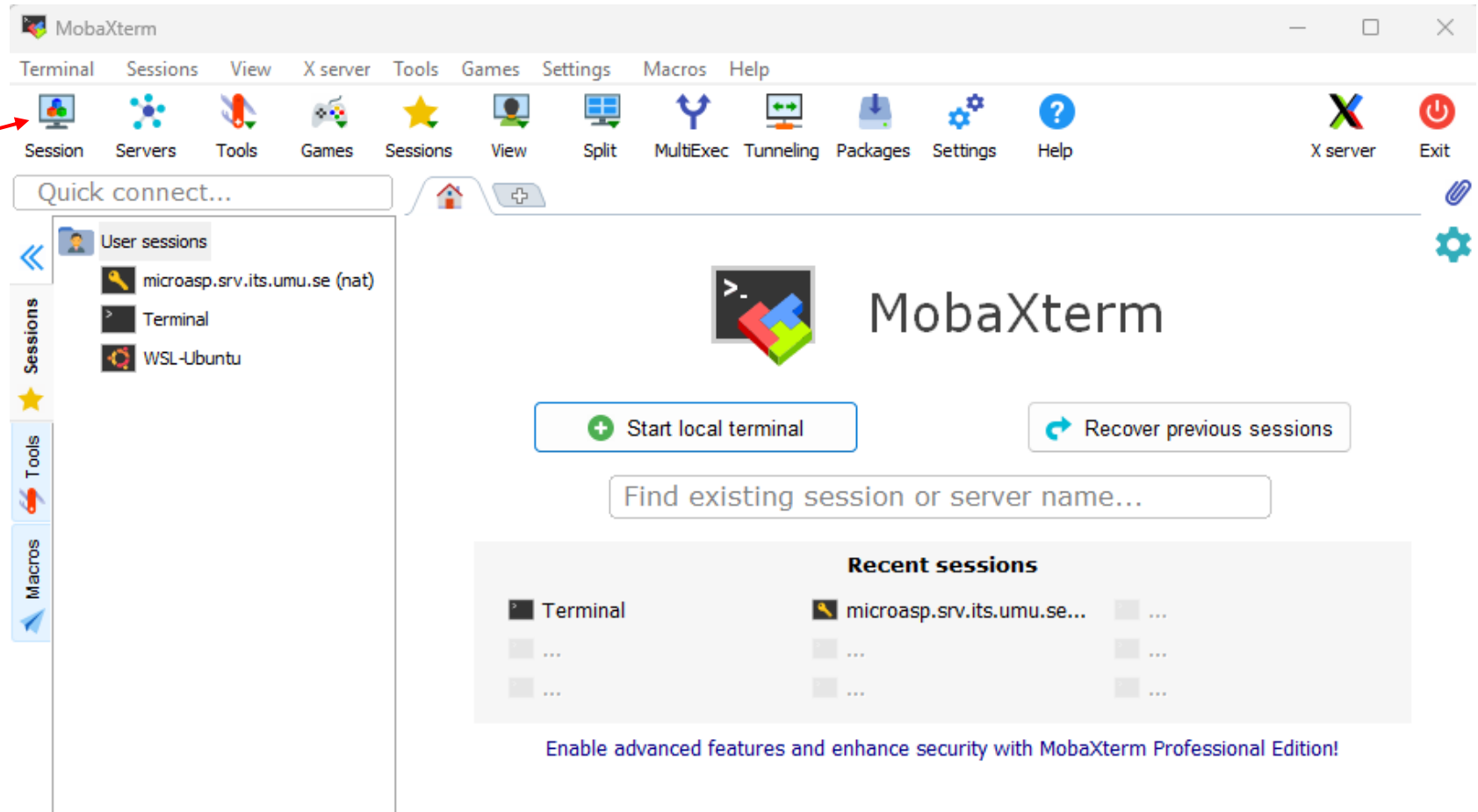
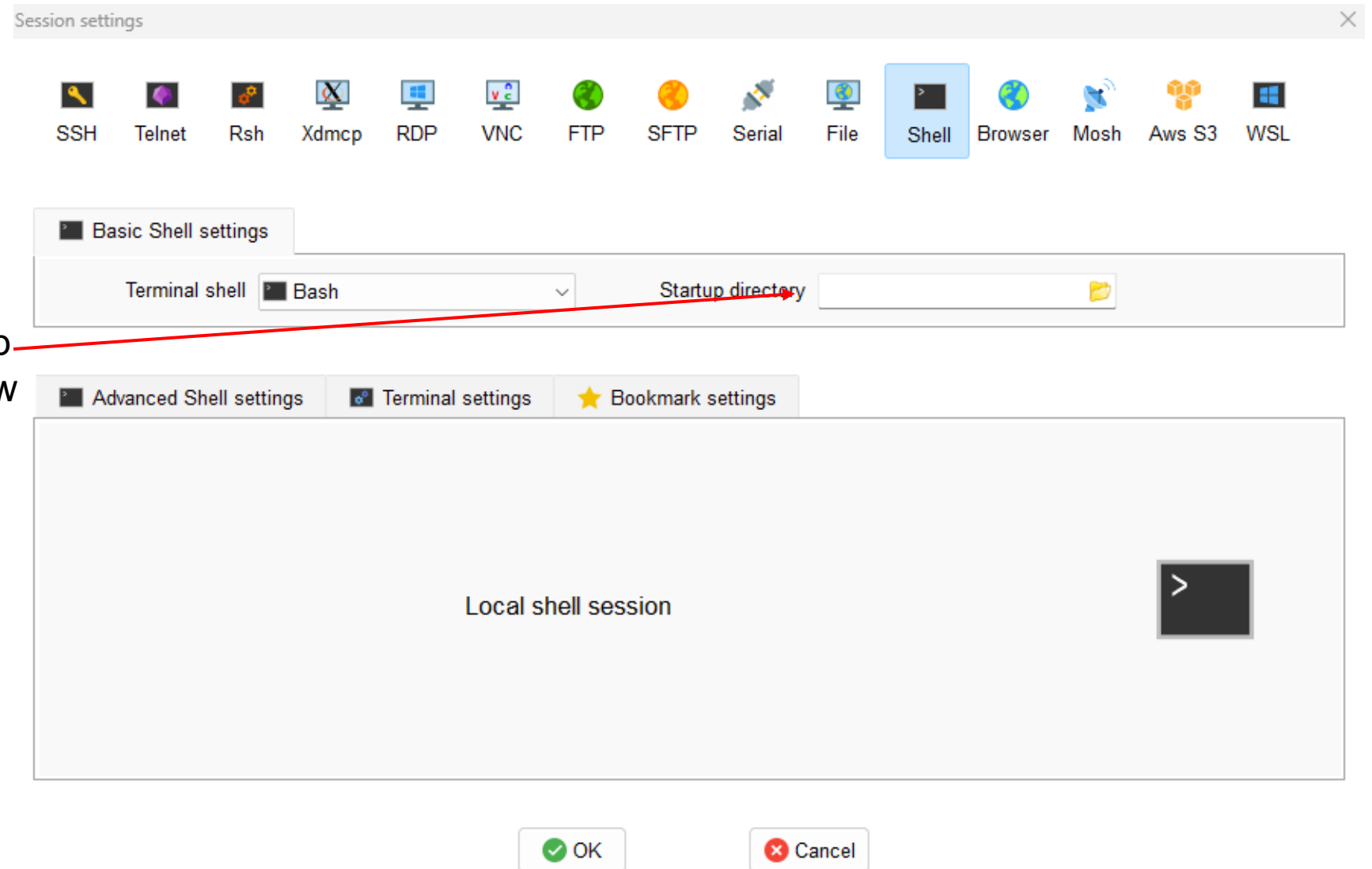- Start a new session

Open a terminal: Windows

- Open MobaXterm

- Start a new session

- Select "shell"

Open a terminal: Windows

- Open MobaXterm

- Start a new session

- Select "shell"

- Select an empty startup directory or make a new one

Umeå Plant Science Centre

Open a terminal: Windows

- Open MobaXterm

- Start a new session

- Select "shell"

- Select an empty startup directory or make a new one

Umeå Plant Science Centre

Open a terminal: Windows

- Open MobaXterm

- Start a new session

- Select "shell"

- Select an empty startup directory or make a new one

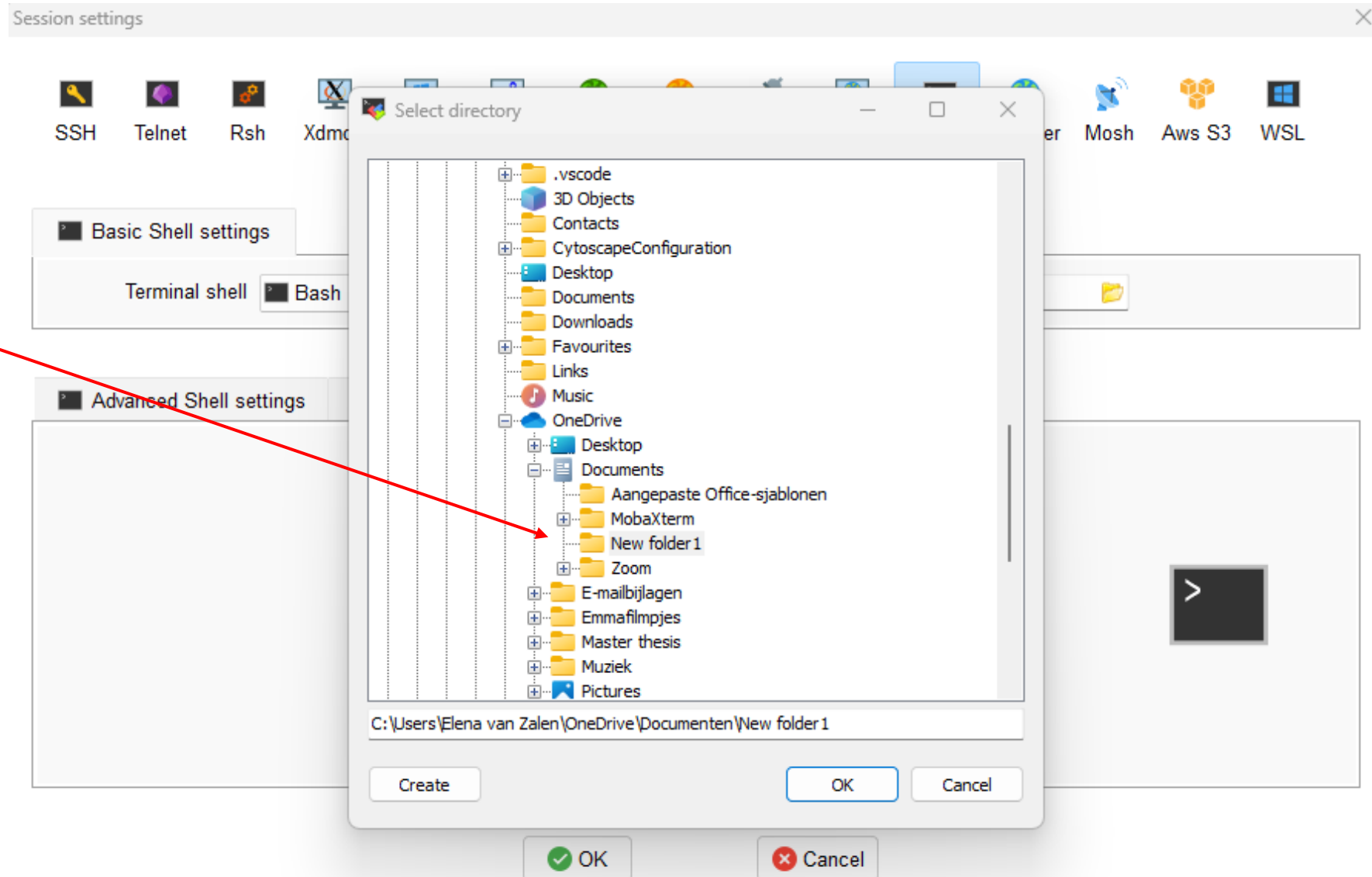- Click "OK" to open a terminal

Umeå Plant Science Centre

Open a terminal: Windows

- Open MobaXterm

- Start a new session

- Select "shell"
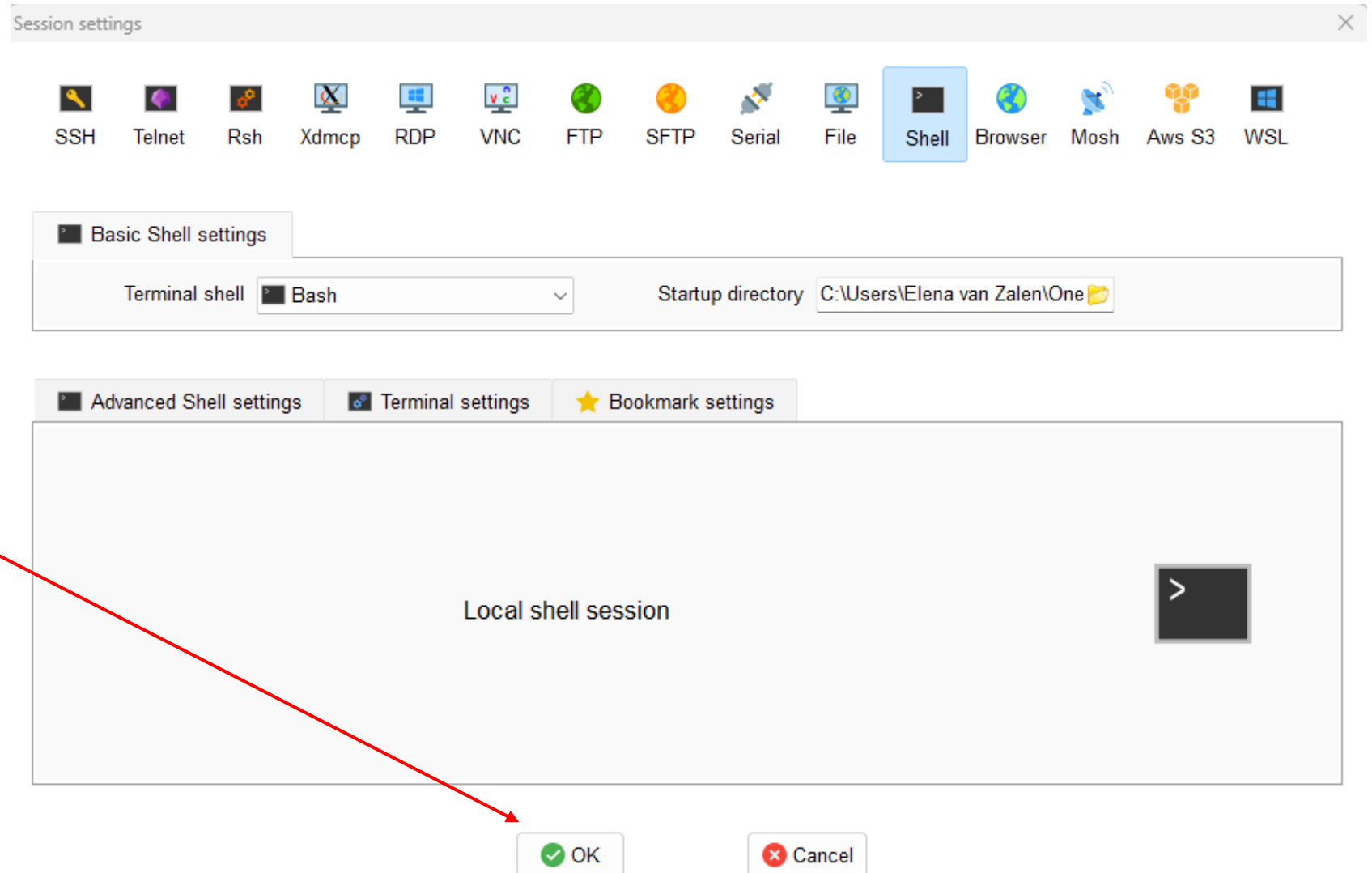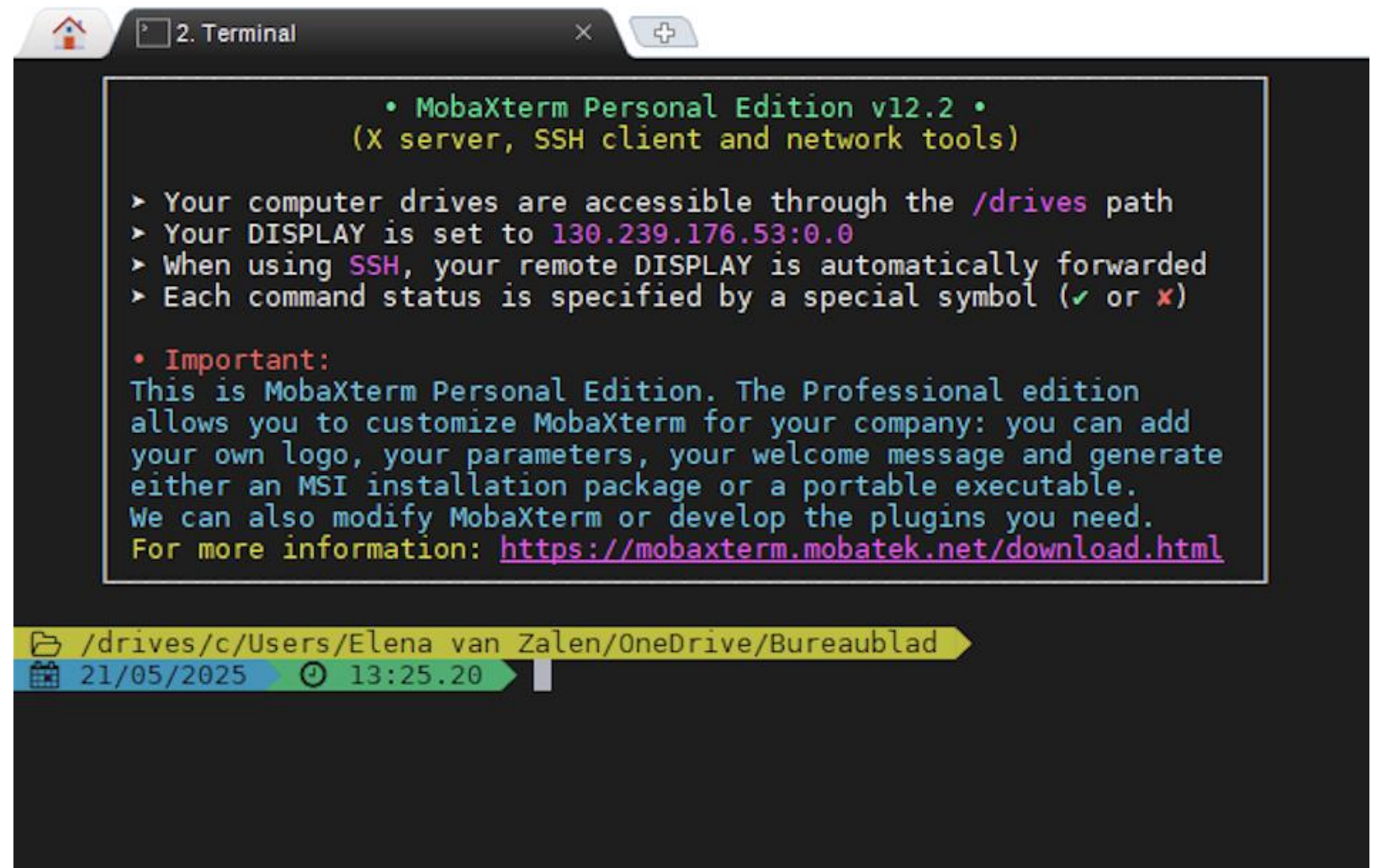
- Select an empty startup directory or make a new one

- Click "OK" to open a terminal

- You should see something like this

Umeå Plant Science Centre

Software carpentry tutorial

Go to this link:

https://swcarpentry.github.io/shell-novice/index.html

Ignore section "Install software" and section "Open a new shell".

Follow the instructions, completing these two sections:

1: Introducing the Shell

2: Navigating Files and Directories

Umeå Plant Science Centre

How to edit files

- In the rest of the lesson, you will edit a variety of files. They are all txt files organized in specific ways, such as **json** files.

- To edit them, use simple text editors and not word, as word tends to add metacharacters that are not correctly interpreted by the softwares we will use.

- Good text editors to use are TextEdit (for Mac) and Notepad or WordPad (for windows).

- A more complex one to use (very good, but with many functionalities that may be confusing) is VScode (https://code.visualstudio.com/download), which works on every system.

Umeå Plant Science Centre

What is nf-core?

From their website: *A global community effort to collect a curated set of open-source analysis pipelines built using Nextflow.*

When starting a new project requiring bioinformatics analyses, it is not a bad idea to check the nf-core website to see if they have a pipeline prepared to run the analysis of the type of data you have:
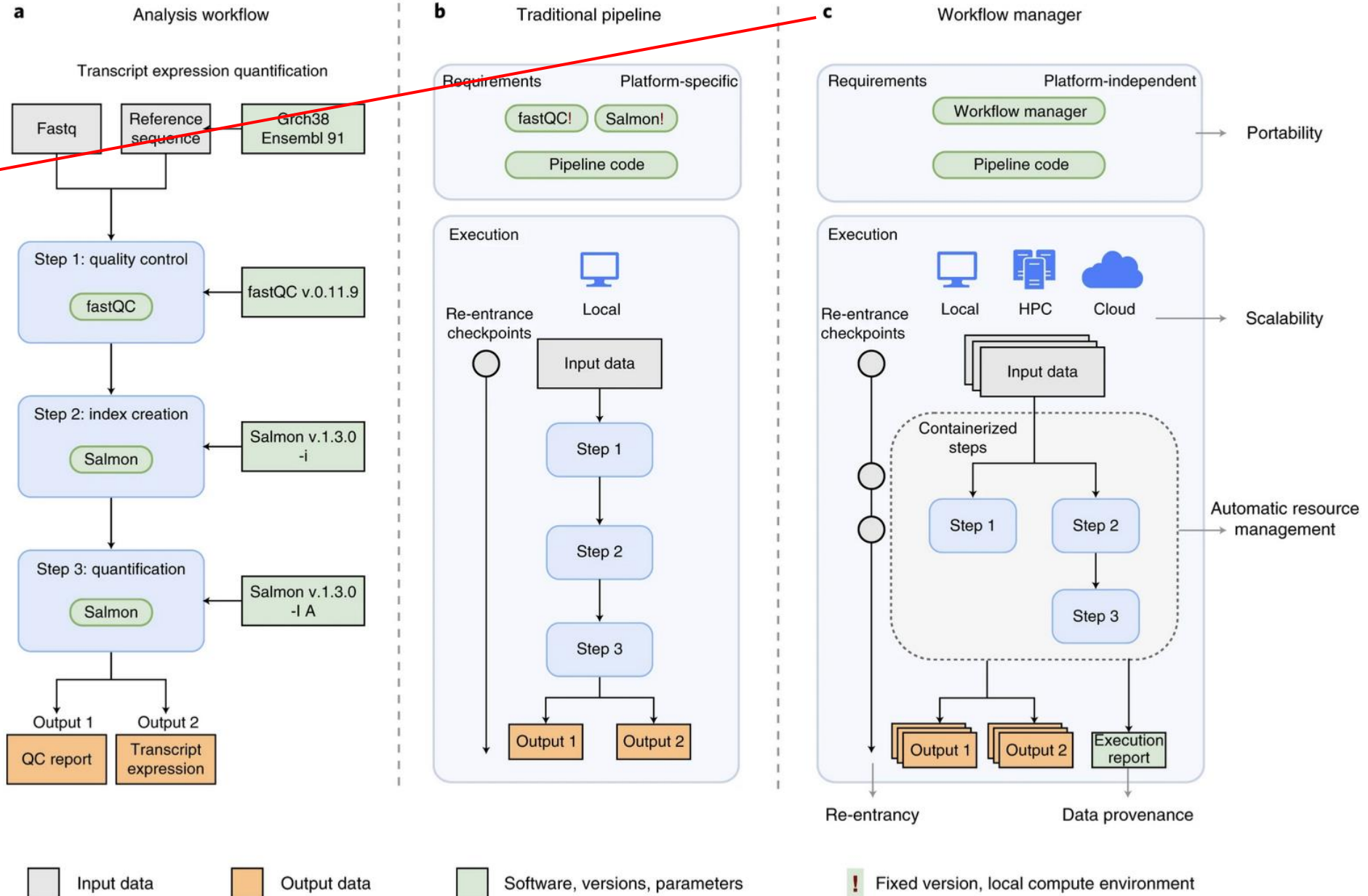
https://nf-co.re/

Visually:

Nextflow is the software allowing this

Nf-core is a repository of nextflow pipelines built for different bioinformatics applications

a Analysis workflow

Transcript expression quantification

- Fastq
- Reference sequence
- Grch38 Ensembl 91

Step 1: quality control
- fastQC
- fastQC v.0.11.9

Step 2: index creation
- Salmon
- Salmon v.1.3.0 -i

Step 3: quantification
- Salmon
- Salmon v.1.3.0 -I A

Output 1 — QC report
Output 2 — Transcript expression

b Traditional pipeline

Requirements — Platform-specific
- fastQC!
- Salmon!
- Pipeline code

Execution
- Re-entrance checkpoints
- Local
- Input data
- Step 1
- Step 2
- Step 3
- Output 1
- Output 2

c Workflow manager

Requirements — Platform-independent
- Workflow manager
- Pipeline code
→ Portability

Execution
- Local, HPC, Cloud → Scalability
- Re-entrance checkpoints
- Input data
- Containerized steps
- Step 1
- Step 2
- Step 3
→ Automatic resource management
- Output 1
- Output 2
- Execution report

Re-entrancy   Data provenance

Legend:
- Input data
- Output data
- Software, versions, parameters
- ! Fixed version, local compute environment

Umeå Plant Science Centre

Running the rnaseq pipeline

- Download my github repository by entering in your shell the following command:

git clone https://github.com/EdoardoPiombo/nf-core-introduction.git

- Follow the instructions in the READme file to do the exercise.

- You have solutions in the "solutions" folder if you are stuck.

- When you are done, feel free to check out the other pipelines and see if there is something that could help you with your work:

https://nf-co.re/pipelines/

Umeå Plant Science Centre