



FORRES: Computational infrastructure to Foster Open and Reproducible REanalysis of Scientific Data

Simon Schwab

Center for Reproducible Science (CRS)

14.12.20

Funding

- For a **computational infrastructure** to support research related to reproducibility and meta-research
- For this service we gratefully receive funding from the StwF.

Linux virtual server

- Virtual Server that supports meta-research
- Accessible for CRS staff, CRS fellows, CRS members
- Hostname: **crs.ebpi.uzh.ch**
- Request account: simon.schwab@uzh.ch
 - Provide your preferred username (usually first name)

Computational infrastructure

Specifically, the server has the following properties:

- Multicore server (CPUs are scalable)
- Ubuntu Linux 20.04 LTS with XFCE desktop environment
- Logical volume management (LVM) → scalable storage space
- R software for statistical computing, RStudio
- Other software possible, on demand (simon.schwab@uzh.ch)
- Online 7x24
- Connect with `ssh` (terminal) or `x2go` (full desktop)
 - x2go client supports Linux, Windows and OS X
 - <https://wiki.x2go.org/doku.php/download:start>
 - in Ubuntu: install package `x2goclient` - X2Go Client application (Qt5)
- UZH VPN required for connection from outside the university

x2go configuration

Set configuration as follows to connect to server:

The screenshot shows the 'Session' tab of the x2go configuration window. The 'Session name' field is set to 'CRS'. Below it is a default seal icon with the text '<< change icon'. The 'Path' field is set to '/'. The 'Server' section contains the following fields: 'Host' (crs.ebpi.uzh.ch), 'Login' (simon), and 'SSH port' (22). There is a field for 'Use RSA/DSA key for ssh connection' with a file selection icon. Below these are several checkboxes: 'Try auto login (via SSH Agent or default SSH key)' (checked), 'Kerberos 5 (GSSAPI) authentication', 'Delegation of GSSAPI credentials to the server', and 'Use Proxy server for SSH connection'. The 'Session type' section has a checkbox for 'Run in X2GoKDrive (experimental)' and a dropdown menu set to 'XFCE' with a 'Command:' field next to it.

The screenshot shows the 'Media' tab of the x2go configuration window. The 'Sound' section has a checkbox for 'Enable sound support' which is unchecked. Below it are radio buttons for 'PulseAudio' (selected) and 'esd'. To the right, there is a checked checkbox for 'Use SSH port forwarding to tunnel sound system connections through firewalls'. At the bottom, there is an unchecked checkbox for 'Client side printing support'.

Virtual server

The screenshot displays a virtual server desktop environment. The top bar shows the user 'simon@crs: ~' and the system name 'X2GO-simon-50-1607957961_stDXFCE_dp32'. The date and time are 'Mon 14 Dec, 17:08' and the user is 'Simon Schwab'.

The main window is RStudio, showing a scatter plot of 'dist' vs 'speed'. The plot shows a positive correlation between speed and distance. The x-axis is labeled 'speed' and ranges from 0 to 25. The y-axis is labeled 'dist' and ranges from 0 to 100. The data points are scattered, showing a general upward trend.

The RStudio interface includes a menu bar (File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help), a toolbar with various icons, and a console window at the bottom. The console shows the following text:

```
13  
14 Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by  
pressing *Ctrl+Alt+I*.  
15  
16 When you save the notebook, an HTML file containing the code and output will be  
saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview  
10.7 Chunk 1
```

The File Manager window shows the home directory '/home/simon/'. It displays a sidebar with 'DEVICES' (File System) and 'PLACES' (simon, Desktop, Trash, Documents, Downloads, Network). The main pane shows a file tree with folders 'Data', 'Desktop', 'Documents', 'Downloads', and 'Public', and files 'R' and 'README'. The status bar indicates '7 items: 289 bytes, Free space: 6.6 GIB'.

The Terminal window shows the following commands and output:

```
simon@crs:~$ ls  
Data Desktop Documents Downloads Public R README  
simon@crs:~$
```

Additional features

- Transfer data with `scp` or `filezilla` (secure port 22)
- Always login to same session and find desktop as you left it
- ...

Use case #1

Meta-research project about publication bias in Cochrane systematic reviews

- Downloading and parsing of >6,000 XML datasets from the Cochrane Database of Systematic Reviews (CDSR)
- Reading and extracting 60,000 HTML tables with study characteristics
- Collecting citation information from Google Scholar for 40,000 intervention studies

Tools used:

- R packages `rvest` and `httr`
- Scripting X Window behavior of Google Chrome with `xdotool` and `bash`
- Software available at project website: <https://osf.io/uv397/>

Use case #2

Post-publication statistical review and method reproducibility project:

“Investigating the appropriate use of ANCOVA in medicine, neuroscience and psychology”

- Downloading and reanalyzing data on the virtual server
- Analysis pipeline that generates a statistical review report

Tools used:

- R Markdown/R Notebooks
- Statistical reports available at project website: <https://osf.io/4mfdh/>