Connecting Data Analysis Platform (DAP)

Document Author: arho.virkki@tyks.fi

Obtaining Access to DAP

Turku Centre for Clinical Informatics at the Hospital District of Southwest Finland offers Data Analytics Platform (DAP) for data-driven research projects.

To use the service, you need to

1. Have a valid research permission (T-number)
2. Have agreed DAP terms and conditions
3. Have obtained credentials for the service

This guide explains how to get credentials to DAP, and the optional step of installing an open-sourced X2Go software package for Linux, Apple or Windows for native remote desktop access.

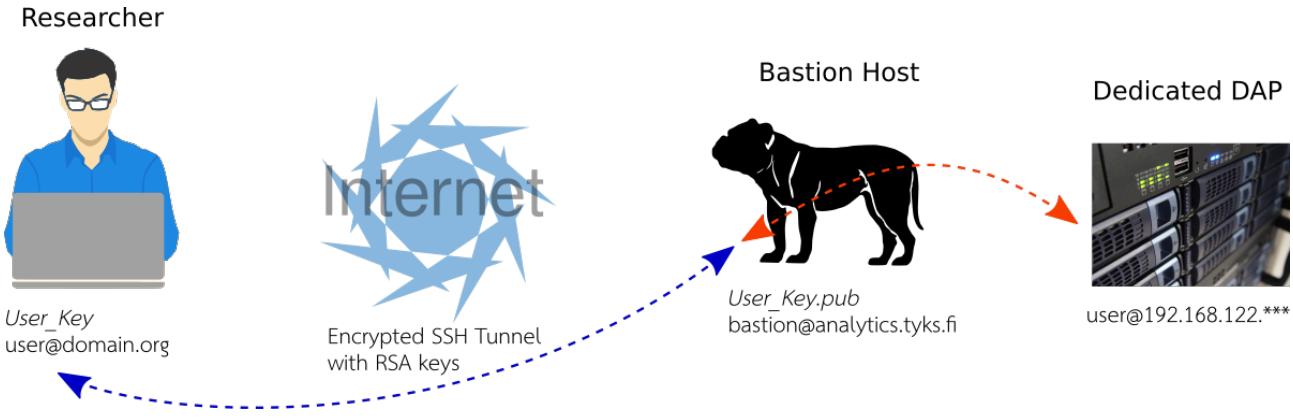


Figure 1:

Architecture. A Schematic illustarion of the Data Analytics Platform (DAP). API end points and web services like "R Studio Server" can be tunneled to researcher's computer through the Secure Shell (SSH) connection, which is a sptandard Internet protocol for encrypted communication.

1. Obtain Key Pair for Asymmetric Cryptography

Visit the CCI office, or generate the files yourself with the instructions below.

On Linux and Apple, the needed tools come bundled. On Windows 10, you need either to enable the Windows Sybsystem for Linux [1], or get a separate tool that installs OpenSSH. The easiest way for Windows users is to install Git [2], which comes with a bundled bash shell and ssh.

When ready, open Terminal.app, Gnome Terminal or Bash Launcher, and generate a key pair with the command

```
ssh-keygen -C "Your Name" -f Your_Name
```

The command generates two files, for example

```
John_Doe
John_Doe.pub
```

Email the public key (with the .pub extension) to ktp@tyks.fi (cc: arho.virkki@tyks.fi). Keep the private key (with no extension) in a safe place. **Do not email the private key to anyone!**

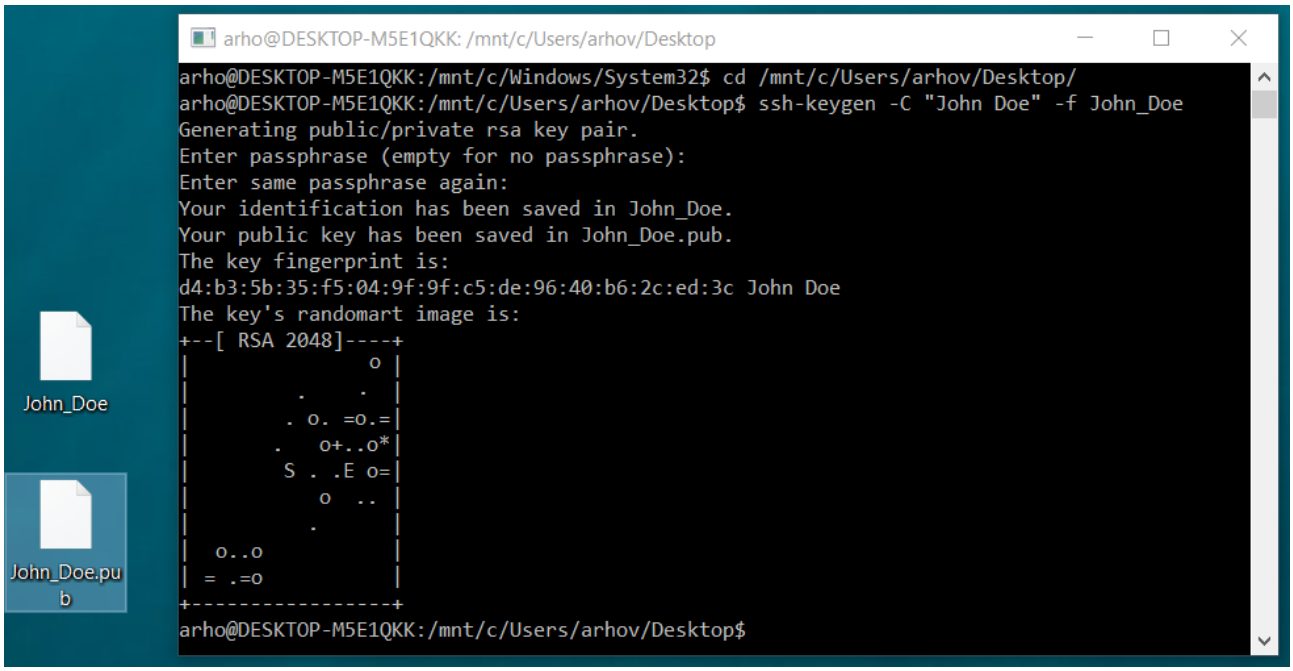


Figure 2:

Example. Generating keys public and private keys.

[1] http://www.howtogeek.com/249966/how-to-install-and-use-the-linux-bash-shell-on-windows-10/
[2] https://git-scm.com/

2. Connect DAP via Command Line

Connection works only after CCI has obtained your public key and added it to the system.

Again, open Terminal.app, Gnome Terminal or Bash Launcher, and issue the command

```
ssh <user>@192.168.122.<your machine> -o \
ProxyCommand="ssh bastion@analytics.tyks.fi -W %h:%p -i <private_key>"
```

where “" denotes the continuation of line. For example,

```
ssh <your user>@192.168.122.<machine_num> -o ProxyCommand="ssh bastion@analytics.tyks.fi -W %h:%p
    -i ~/c/files/KTP_Key"
```

You should be able to log in with the correct password.

Using services on the DAP is now easy. For example, you can tunnel R Studio Server through the connection by adding its port as an option in the above command,

```
-L 8787:localhost:8787
```

Now, R Studio Server can be reached by pointing a browser to address http://localhost:8787, as shown in the image below.

Example. Using R Studio Server on DAP.

3. Connect DAP Using X2Go Remote Desktop

Download and install a copy of X2Go from http://wiki.x2go.org. This will provide you a full-blown graphical desktop to the remote server.

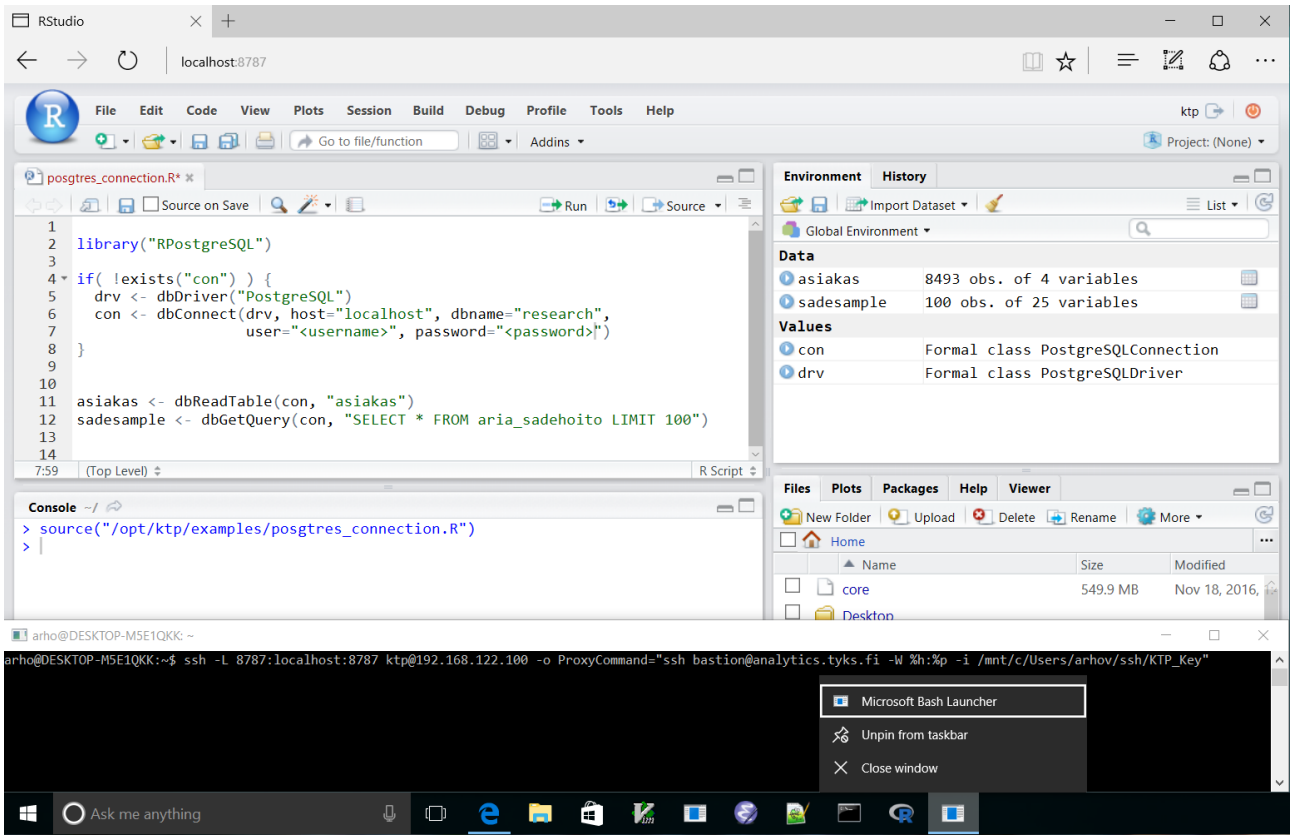


Figure 3:

The settings must be chosen exactly as in the image below.

Note! Some recent **OS X** versions do have trouble in connecting through the bastion SSH proxy. Stable and tested versions are available at: http://cci.vsshp.fi/public/x2go/apple.

- Host: the IP of your machine, e.g. 192.168.122.100
- Login: You account name, e.g. johndoe
- [x] Use Proxy server for SSH connection
- Type: SSH
- Login: bastion
- Host: analytics.tyks.fi
- RSA/DSA key: Full path to your *private* key
- Session type: XFCE

Example. Connection parameters for X2Go.

Known limitations

- On some versions of Windows, X2Go does not suppors switching between full screen and windowed mode during a single session, or does not support full screen mode at all.
- If the path to the RSA/DSA private key is wrong, the software will close with no error message.

Accessing the Data

In most cases, the data is stored in PosgreSQL database engine in a database named *research*. A ready-made ordinary user *analyst* (with the same default password) can access this database from the virtual machine's terminal with the command

```
PGPASSWORD=analyst psql -U analyst -d research -h localhost
```

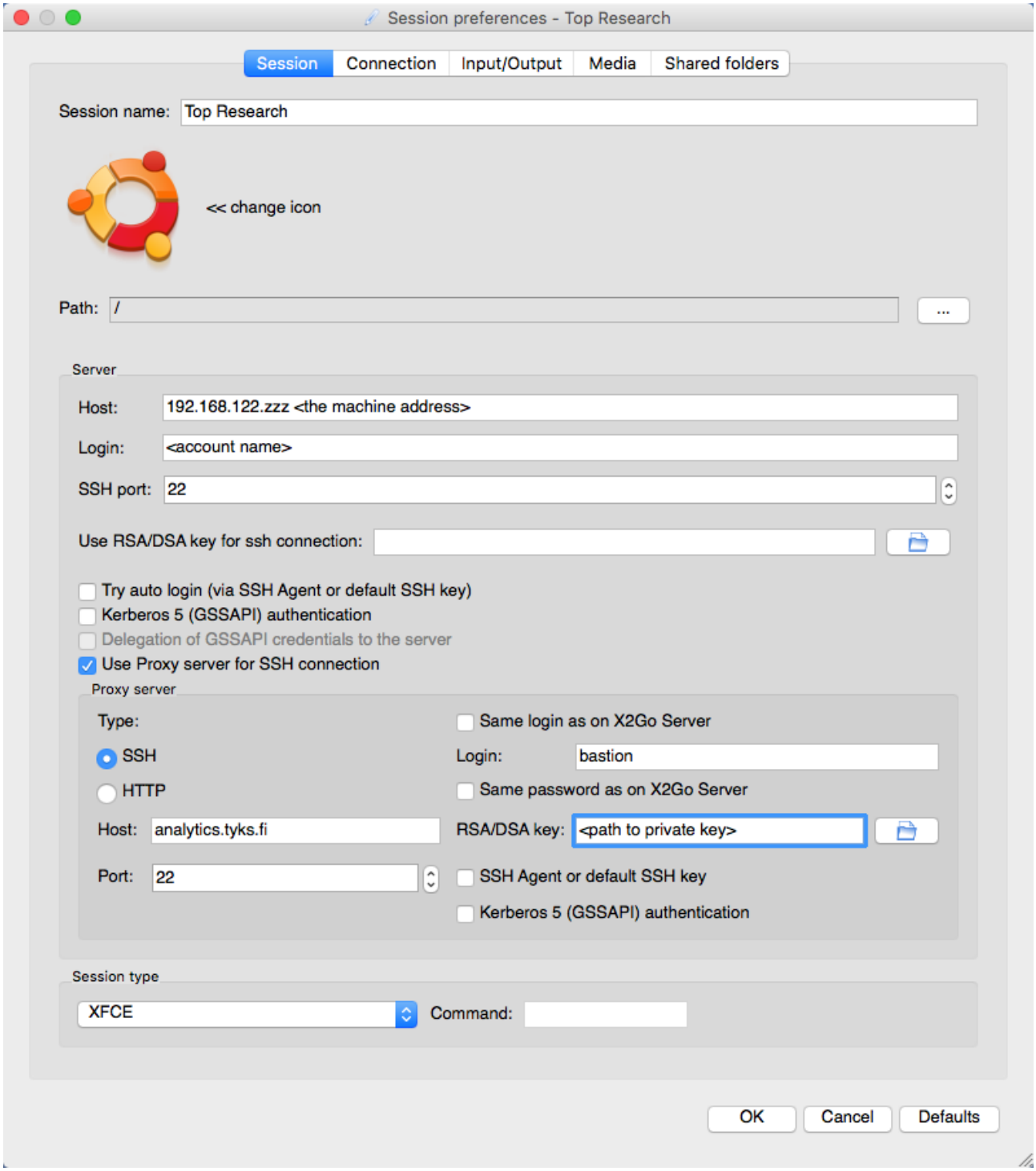

Session preferences - Top Research
Session
Connection
Input/Output
Media
Shared folders
Session name: Top Research
<< change icon
Path: /
...
Server
Host: 192.168.122.zzz <the machine address>
Login: <account name>
SSH port: 22
Use RSA/DSA key for ssh connection:
Try auto login (via SSH Agent or default SSH key)
Kerberos 5 (GSSAPI) authentication
Delegation of GSSAPI credentials to the server
Use Proxy server for SSH connection
Proxy server
Type:
SSH
HTTP
Host: analytics.tyks.fi
Port: 22
Same login as on X2Go Server
Login: bastion
Same password as on X2Go Server
RSA/DSA key: <path to private key>
SSH Agent or default SSH key
Kerberos 5 (GSSAPI) authentication
Session type
XFCE
Command:
OK
Cancel
Defaults

Figure 4:

The data resides in the *data* schema. The tables in this schema can be listed as follows:

```
research=> \dt data.
                   List of relations
 Schema |              Name              | Type  | Owner
--------+--------------------------------+-------+-------
 data   | asiakas                        | table | ktp
 data   | diagnoosi                      | table | ktp
 data   | laake_maarays                  | table | ktp
 data   | labrat                         | table | ktp
 data   | leikkaus_opera                 | table | ktp
 data   | leikkaus_toti                  | table | ktp
 data   | oberon_toimenpiteet            | table | ktp
 data   | palvelu                        | table | ktp
 data   | radut                          | table | ktp
 data   | reseptit                       | table | ktp
 data   | yhdistetty                     | table | ktp
(11 rows)
```

Uploading Own Data

Option 1.

Use *X2Go Shared folders* feature to copy data back and forth between the computing environment.

Option 2.

Use command line *scp* (With Linux, OS X, and Microsoft's Windows 10 Subsystem for Linux)

Examples:

```
# Open a tunnel to the remote machine at local port 2222.
ssh -N -L 2222:192.168.122.100:22 bastion@analytics.tyks.fi -i <private_key_file> &

# Copy the file(s) to DAP Desktop folder
scp -P 2222 <my_files> <username>@localhost:~/Desktop/
```

Option 3.

Use dedicated secure copy software like WinSCP on Windows which can tunnel connections through the bastion host.

Installing WinCSP (for Windows)

Download a recent copy of WinSCP from https://winscp.net/eng/download.php and run the installation package. (In case you do not have administrator rights, choose portable executables. They are just launched directly and not installed.)

Configure site. Set host name to your DAP IP (192.168.122.<number>) and user name and password to your personal credentials. Then click the "Advanced…" button.

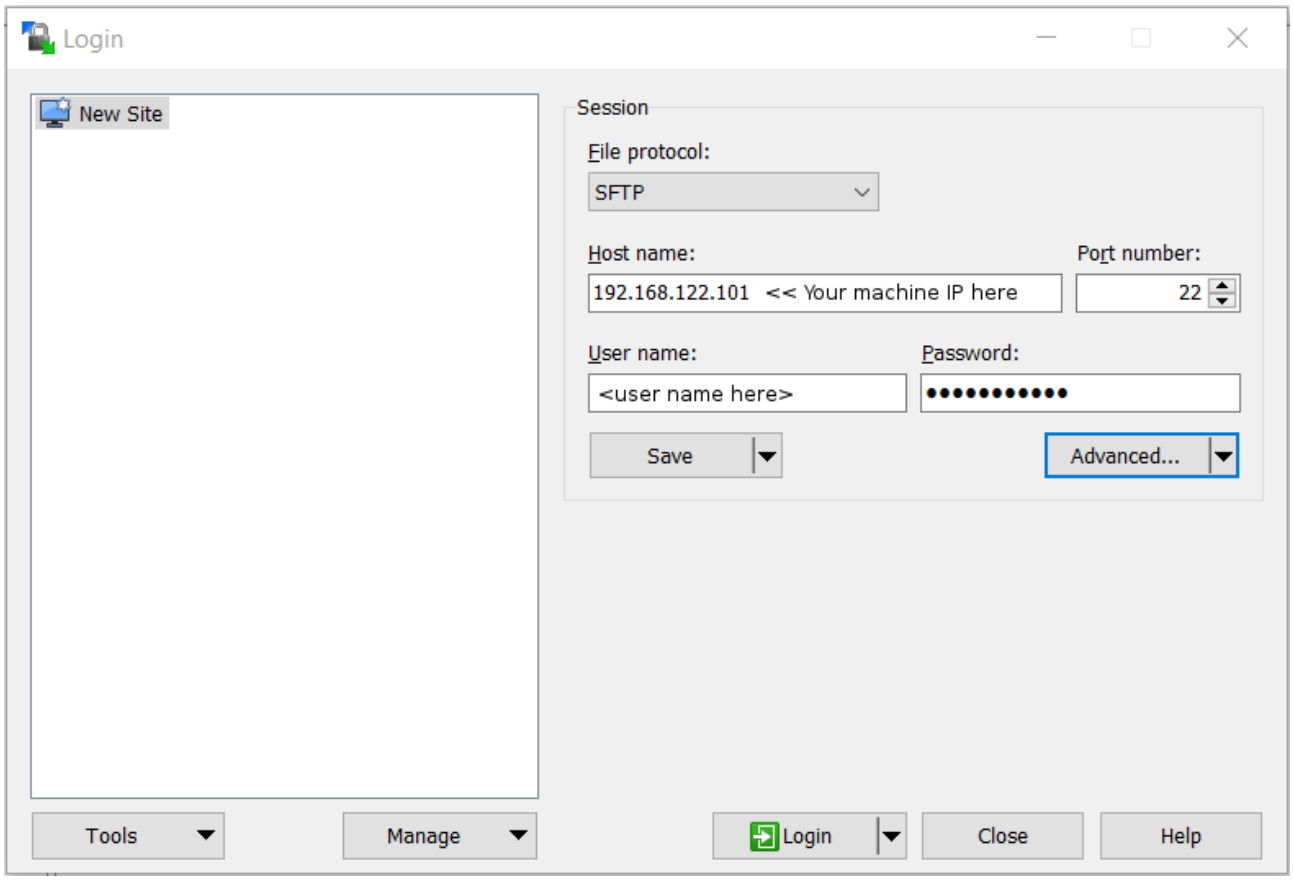
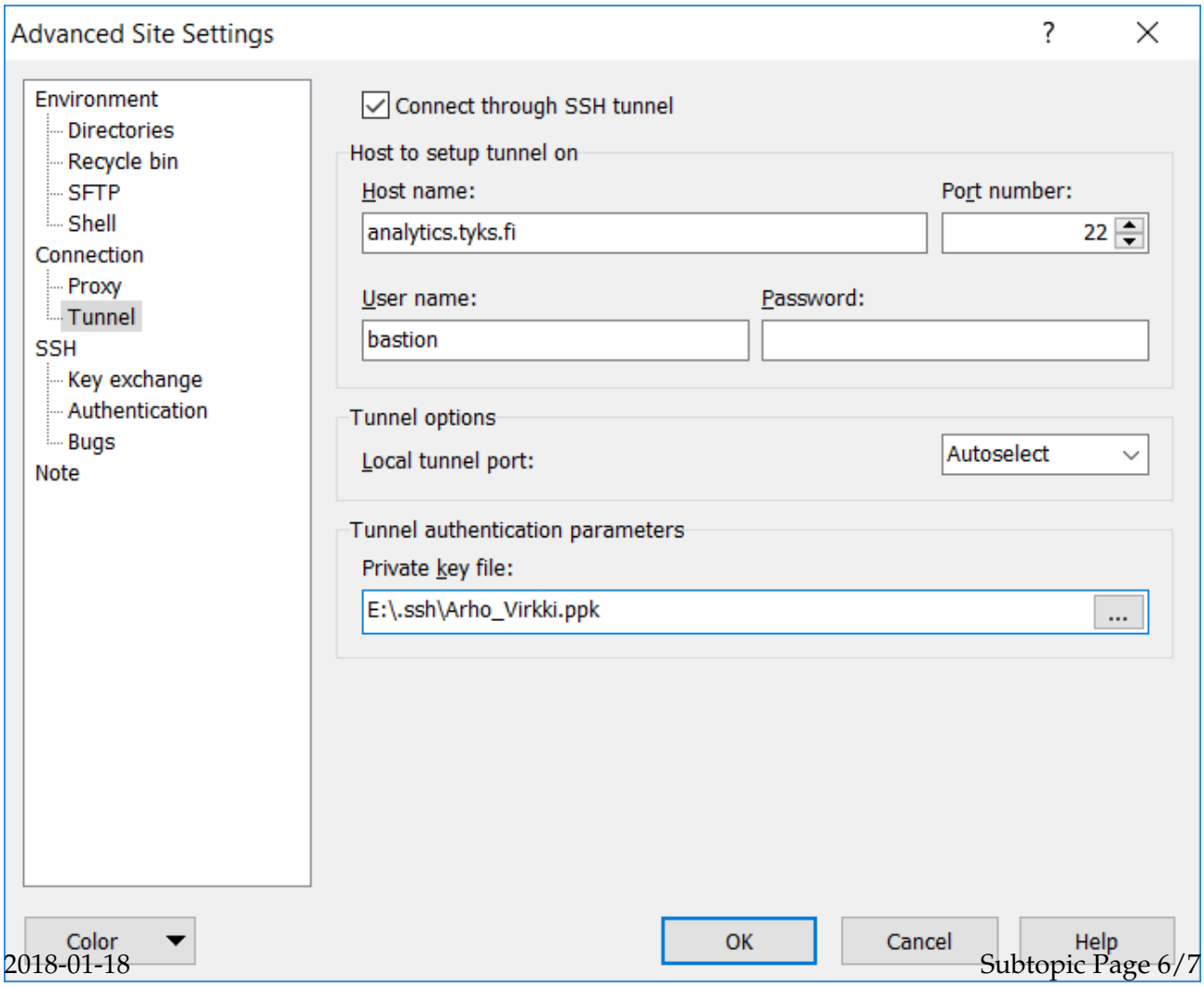

Login
New Site
Session
File protocol:
SFTP
Host name:
Port number:
192.168.122.101 << Your machine IP here
22
User name:
Password:
<user name here>
Save
Advanced...
Tools
Manage
Login
Close
Help

Figure 5:


Advanced Site Settings
Environment
Directories
Recycle bin
SFTP
Shell
Connection
Proxy
Tunnel
SSH
Key exchange
Authentication
Bugs
Note
Connect through SSH tunnel
Host to setup tunnel on
Host name:
Port number:
analytics.tyks.fi
22
User name:
Password:
bastion
Tunnel options
Local tunnel port:
Autoselect
Tunnel authentication parameters
Private key file:
E:\.ssh\Arho_Virkki.ppk
...
Color
OK
Cancel
Help

Set up tunnel. Select Connection -> Tunnel,and [x] Connect through SSH tunnel. Host name must be **analytics.tyks.fi** and user **bastion**. Finally choose your private key file. When given an OpenSSH privater key, WinSCP offers to convert it into (Windows-specific) Putty format.

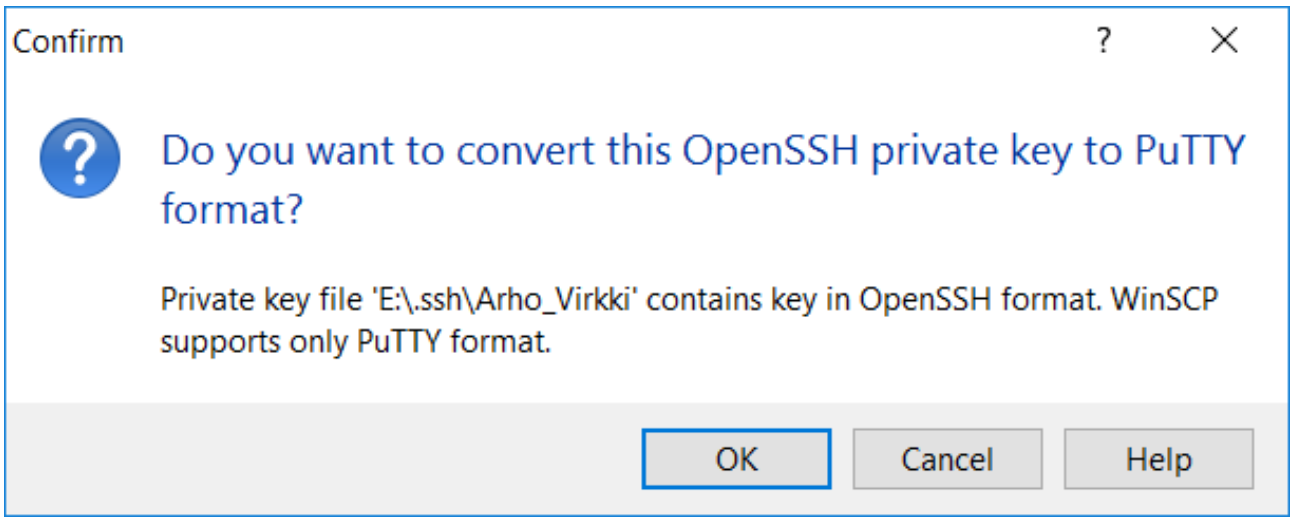


Figure 6:

Key conversion. OpenSSH private key can be converted into Putty format and saved.

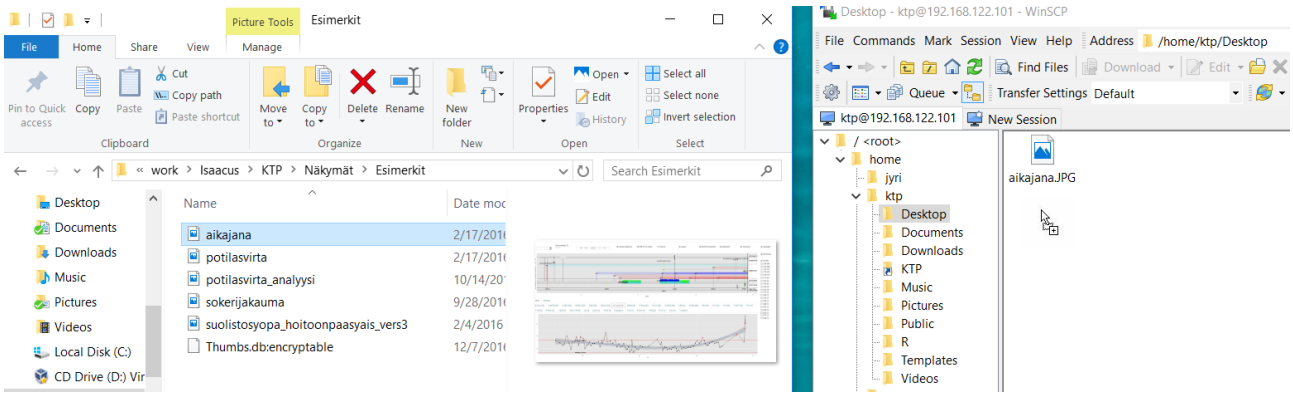


Figure 7:

File copying. Files can be copied with drag and drop.